

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> A01H 4/00, 5/00, C12N 15/29, 15/31, 15/65, 15/67, 15/82, 15/84		<b>A1</b>	<b>(11) International Publication Number:</b> WO 00/07431 <b>(43) International Publication Date:</b> 17 February 2000 (17.02.00)															
<b>(21) International Application Number:</b> PCT/US99/17806 <b>(22) International Filing Date:</b> 3 August 1999 (03.08.99)		<b>(74) Agents:</b> RIGAUT, Kathleen, D. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103 (US).																
<b>(30) Priority Data:</b> <table border="0"><tr><td>60/095,163</td><td>3 August 1998 (03.08.98)</td><td>US</td></tr><tr><td>60/095,167</td><td>3 August 1998 (03.08.98)</td><td>US</td></tr><tr><td>60/112,257</td><td>15 December 1998 (15.12.98)</td><td>US</td></tr><tr><td>60/131,611</td><td>29 April 1999 (29.04.99)</td><td>US</td></tr><tr><td>60/138,764</td><td>11 June 1999 (11.06.99)</td><td>US</td></tr></table>		60/095,163	3 August 1998 (03.08.98)	US	60/095,167	3 August 1998 (03.08.98)	US	60/112,257	15 December 1998 (15.12.98)	US	60/131,611	29 April 1999 (29.04.99)	US	60/138,764	11 June 1999 (11.06.99)	US	<b>(81) Designated States:</b> AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
60/095,163	3 August 1998 (03.08.98)	US																
60/095,167	3 August 1998 (03.08.98)	US																
60/112,257	15 December 1998 (15.12.98)	US																
60/131,611	29 April 1999 (29.04.99)	US																
60/138,764	11 June 1999 (11.06.99)	US																
<b>(71) Applicant (for all designated States except US):</b> RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY [US/US]; Old Queens, Somerset Street, New Brunswick, NJ 08903 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> MALIGA, Pal [HU/US]; 72 Yorktown Road, East Brunswick, NJ 08816 (US). KURODA, Hiroshi [JP/US]; Apartment 2C, 2 Redcliffe Avenue, Highland Park, NJ 08904 (US). KHAN, Muhammad, Sarwar [PK/US]; Russels Apartment, 761 Bevier Road, Piscataway, NJ 08854 (US).		<b>Published</b> With international search report.																
<b>(54) Title:</b> TRANSLATION CONTROL ELEMENTS FOR HIGH-LEVEL PROTEIN EXPRESSION IN THE PLASTIDS OF HIGHER PLANTS AND METHODS OF USE THEREOF																		
<b>(57) Abstract</b> <p>DNA constructs containing translational control elements are provided. These 5' regulatory segments facilitate high level expression of transgenes introduced into the plastids of higher plants.</p>																		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**Translation Control Elements for High-level Protein  
Expression in the Plastids of Higher Plants and Methods  
of Use Thereof**

5

10           This application claims priority from United States  
Provisional Applications 60/095,163, filed August 3,  
1998, 60/112,257, filed December 15, 1998, 60/095,167  
filed August 3, 1998, 60/131,611, filed April 29, 1999  
and 60/138,764, filed June 11, 1999 under 35 U.S.C.  
15 §119(e). The entire disclosures of each of the  
foregoing are incorporated by reference herein.

          Pursuant to 35 U.S.C. §202(c) it is acknowledged  
that the U.S. Government has certain rights in the  
20 invention described herein, which was made in part with  
funds from the National Science Foundation, Grant Number  
MCB-96-30763.

**FIELD OF THE INVENTION**

25           This invention relates to the fields of transgenic  
plants and molecular biology. More specifically, the  
invention provides vectors targeting the plastid genome  
which contain translation control elements facilitating  
high levels of protein expression in the plastids of  
30 higher plants. Both monocots and dicots are  
successfully transformed with the DNA constructs  
provided herein.

**BACKGROUND OF THE INVENTION**

35           Several publications are referenced in this  
application in order to more fully describe the state of

the art to which this invention pertains. The disclosure of each of these publications is incorporated by reference herein.

5       The chloroplasts of higher plants accumulate individual components of the photosynthetic machinery as a relatively large fraction of total cellular protein. The best example is the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco) involved in CO<sub>2</sub> fixation which can make up 65% of the total leaf protein (Ellis, 10       R.J. 1979). Because of the potentially attainable high protein levels, there is significant interest in exploring chloroplasts as an alternative system for protein expression. To date, protein levels expressed from transgenes in chloroplasts are below the levels of 15       highly-expressed chloroplast genes. Highest levels reported thus far in leaves are as follows: 1% of neomycin phosphotransferase (Carrer et al., 1993); 2.5%  $\beta$ -glucuronidase (Staub and Maliga, 1993) and 3-5% of *Bacillus thuringiensis* (Bt) crystal toxins (McBride et 20       al., 1995). An alternative system, based on a nuclear-encoded, plastid-targeted T7 RNA polymerase may offer higher levels of protein expression (McBride et al., 1994), although this yield may come at a price.

25       In bacteria, the rate limiting step of protein synthesis is usually the initiation of translation, involving the binding of the initiator tRNA (formyl-methionyl-tRNA<sub>f</sub>) and mRNA to the 70S ribosome, recognition of the initiator codon, and the precise phasing of the reading frame of the mRNA. Translation 30       initiation depends on three initiation factors (IF1, IF2, IF3) and requires GTP. The 30S subunit is guided to the initiation codon by RNA-RNA base pairing between the 3' of the 16S rRNA and the mRNA ribosome binding site,

or Shine-Dalgarno (SD) sequence, located about 10 nucleotides upstream of the translation initiation codon (Voorma, 1996). RNA-RNA interaction between the "downstream box" (DB), a 15 nt sequence downstream of the AUG translational initiation codon and complementary sequences in the 16S rRNA 3' sequence or anti-downstream box (ADB; nucleotide positions 1469-1483) may also facilitate loading of the mRNA onto the 30S ribosome subunit (Sprengart et al., 1996). In addition, specific protein-RNA interactions may also facilitate translation initiation (Voorma, 1996).

Key components of the prokaryotic translation machinery have been identified in plastids, including homologues of the bacterial IF1, IF2 and IF3 initiation factors and an S1-like ribosomal protein (Stern et al., 1997). Most plastid mRNAs (92%) contain a ribosome binding site or SD sequence: GGAGG, or its truncated tri- or tetranucleotide variant. This sequence is similar to the bacterial SD consensus 5'-UAAGGAGGUGA-3' (Voorma, 1996). High level expression of foreign genes of interest in the plastids of higher plants is extremely desirable. The present invention provides novel genetic translational control elements for use in plastid transformation vectors. Incorporation of these elements into such vectors results in protein expression levels comparable to those observed for highly expressed chloroplast genes in both monocots and dicots.

#### SUMMARY OF THE INVENTION

5' genetic regulatory regions contain promoters with distinct DNA sequence information which facilitates recognition by the RNA polymerase and translational control elements which facilitate translation. Both of

these components act together to drive gene expression.

In accordance with the present invention, chimeric 5' regulatory regions have been constructed which incorporate translation control elements. Incorporation of these chimeric 5' regulatory regions into plastid transforming vectors followed by transformation of target plant cells gives rise to dramatically enhanced levels of protein expression. These chimeric 5' regulatory regions may be used to advantage to express foreign genes of interest in a wide range of plant tissues. It is an object of the present invention to provide DNA constructs and methods for stably transforming plastids of multicellular plants containing such promoters.

In one embodiment of the invention recombinant DNA constructs for expressing at least one heterologous protein in the plastids of higher plants are provided. The constructs comprise a 5' regulatory region which includes a promoter element, a leader sequence and a downstream box element operably linked to a coding region of said at least one heterologous protein. The chimeric regulatory region acts to enhance translational efficiency of an mRNA molecule encoded by said DNA construct. Vectors comprising the DNA constructs are also contemplated in the present invention. Exemplary DNA constructs of the invention include the following chimeric regulatory regions: PrnnLatpB+DBwt, PrnnLatpB-DB, PrnnLatpB+DBm, PrnnLclpP+DBwt, PrnnlclpP-DB, PrnnLrbcL+DBwt, PrnnLrbcL-DB, PrnnLrbcL+DBm, PrnnLpsbB+DBwt, PrnnLpsbB-DB, PrnnLpsbA+DBwt, PrnnLpsbA-DB, PrnnLpsbA-DB(+GC), PrnnLT7g10+DB/Ec, PrnnLT7g10+DB/pt, and PrnnLT7g10-DB. Downstream box sequences preferred for use in the constructs of the

invention have the following sequences:

5'TCCAGTCACTAGCCCTGCCTTCGGCA'3 and 5'CCCAGTCATGAATCACA  
AAGTGGTAA'3.

5 The 5' regulatory segments of the invention have  
been successfully employed to drive the expression of  
the bar gene from *S. hydroscopicus* in the plastids of  
higher plants. Synthetic bar genes have also been  
generated and expressed using the DNA constructs of the  
present invention. These constructs have been  
10 engineered to maximize transgene containment in plastids  
by incorporating rare codons into the coding region that  
are not preferred for protein translation in  
microorganisms and fungi.

In yet another embodiment of the invention, at  
15 least one fusion protein is produced utilizing the DNA  
constructs of the invention. An exemplary fusion  
protein has a first and second coding region operably  
linked to the 5' regulatory regions described herein  
such that production of said fusion protein is regulated  
20 by said 5' regulatory region. In one embodiment the  
first coding region encodes a selectable marker gene and  
the second coding region encodes a fluorescent molecule  
to facilitate visualization of transformed plant cells.  
Vectors comprising a DNA construct encoding such a  
25 fusion protein are also within the scope of the present  
invention. An exemplary fusion protein consists an aadA  
coding region operably linked to a green fluorescent  
protein coding region. These moieties may be linked by  
peptide linkers such as ELVEGKLELVEGLKVA and  
30 ELAVEGKLEVA.

Plasmids for transforming the plastids of higher  
plants, are also included in the present invention.  
Exemplary plasmids are selected from the group

consisting of pHK30(B), pHK31(B), pHK60, pHK32(B),  
pHK33(B), pHK34(A), pHK35(A), pHK64(A), pHK36(A),  
pHK37(A), pHK38(A), pHK39(A), pHK40(A), pHK41(A),  
pHK42(A), pHK43(A), pMSK56, pMSK57, pMSK48, pMSK49,  
5 pMSK35, pMSK53 and pMSK54.

Transgenic plants, both monocots and dicots  
harboring the plasmids set forth above are also  
contemplated to be within the scope of the invention.

In yet another embodiment of the invention, methods  
10 are provided for producing transplastomic monocots. One  
method comprises a) obtaining embryogenic cells;  
b) exposing said cells to a heterologous DNA molecule  
under conditions whereby said DNA enters the plastids of  
said cells, said heterologous DNA molecule encoding at  
15 least one exogenous protein, said at least one exogenous  
protein encoding a selectable marker; c) applying a  
selection agent to said cells to facilitate sorting of  
untransformed plastids from transformed plastids, said  
cells containing transformed plastids surviving and  
20 dividing in the presence of said selection agent; d)  
transferring said surviving cells to selective media to  
promote plant regeneration and shoot growth; and e)  
rooting said shoots, thereby producing transplastomic  
monocot plants. The heterologous DNA molecule may be  
25 introduced into the plant cell via a process selected  
from the group consisting of biolistic bombardment,  
Agrobacterium-mediated transformation, microinjection  
and electroporation. In one embodiment of the above  
described method, protoplasts are obtained from the  
30 embryogenic cells and the heterologous DNA molecule is  
delivered to said protoplasts by exposure to  
polyethylene glycol. Suitable selection agents for the  
practice of the methods of the invention are



streptomycin, and paromomycin. Monocot plants which may be transformed using the methods of the invention include but are not limited to maize, millet, sorghum, sugar cane, rice, wheat, barley, oat, rye, and turf  
5 grass.

In a preferred embodiment a method for producing transplastomic rice plants is provided. This method entails the following steps: a) obtaining embryogenic calli; b) inducing proliferation of calli on modified  
10 CIM medium; c) obtaining embryogenic cell suspensions of said proliferating calli in liquid AA medium; d) bombarding said embryogenic cells with microprojectiles coated with plasmid DNA; e) transferring said bombarded cells to selective liquid  
15 AA medium; f) transferring said cells surviving in AA medium to selective RRM regeneration medium for a time period sufficient for green shoots to appear; and g) rooting said shoots in a selective MS salt medium.

Plasmids suitable for transforming rice as set forth above include pMSK35 and pMSK53, pMSK54 and  
20 pMSK49. Transplastomic rice plants so produced are also contemplated to be within the scope of the invention.

In yet a final embodiment of the invention methods for containing transgenes in transformed plants are  
25 provided. An exemplary method includes the following steps: a) determining the codon usage in said plant to be transformed and in microbes found in association with said plant; and b) genetically engineering said transgene sequence via the introduction of rare  
30 microbial codons to abrogate expression of said transgene in said plant associated microbe. In an exemplary embodiment of the method described immediately above the transgene is a bar gene and said rare codons

are arginine encoding codons selected from the group consisting of AGA and AGG, and transgene is not expressed in *E.coli*.

5           The following definitions will facilitate the understanding of the subject matter of the present invention:

Heteroplastomic: refers to the presence of a mixed population of different plastid genomes within a single  
10   plastid or in a population of plastids contained in plant cells or tissues.

Homoplastomic: refers to a pure population of plastid genomes, either within a plastid or within a population contained in plant cells and tissues.  
15   Homoplastomic plastids, cells or tissues are genetically stable because they contain only one type of plastid genome. Hence, they remain homoplastomic even after the selection pressure has been removed, and selfed progeny are also homoplastomic. For purposes of the present  
20   invention, heteroplastomic populations of genomes that are functionally homoplastomic (i.e., contain only minor populations of wild-type DNA or transformed genomes with sequence variations) may be referred to herein as "functionally homoplastomic" or "substantially  
25   homoplastomic." These types of cells or tissues can be readily purified to a homoplastomic state by continued selection.

Plastome: the genome of a plastid.

Transplastome: a transformed plastid genome.

30   Transformation of plastids: stable integration of transforming DNA into the plastid genome that is transmitted to the seed progeny of plants containing the transformed plastids.

**Selectable marker gene:** the term "selectable marker gene" refers to a gene that upon expression confers a selective advantage to the plastids and a phenotype by which successfully transformed plastids or cells or tissues carrying the transformed plastid can be identified.

**Transforming DNA:** refers to homologous DNA, or heterologous DNA flanked by homologous DNA, which when introduced into plastids becomes part of the plastid genome by homologous recombination.

**Operably linked:** refers to two different regions or two separate genes spliced together in a construct such that both regions will function to promote gene expression and/or protein translation.

The detailed description as follows provides examples of preferred methods for making and using the DNA constructs of the present invention and for practicing the methods of the invention. Any molecular cloning and recombinant DNA techniques not specifically described are carried out by standard methods, as generally set forth, for example in Sambrook et al., "DNA Cloning, A Laboratory Manual," Cold Spring Harbor Laboratory, 1989.

## BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1A.** Plastid mRNAs and the small (16S) ribosomal RNA contain complementary sequences downstream of AUG implicating interactions between mRNA and 16S rRNA during translation initiation in plastids. Proposed model is based on data in *E. coli* (Sprengart et al., 1996); for sequence of 16S rRNA see ref. (Shinozaki et al., 1986b). SD, Shine-Dalgarno sequence; ASD, anti SD region; DB, downstream box; ADB, anti DB region. Watson-

Crick (line) and G-U (closed circle) pairing are marked.

**Figure 1B.** Sequence of the anti-downstream-box regions (ADB sequence underlined) of the 16S rRNA in plastids (pt; this application) and in *E. coli* (Ec; Sprengart et al., 1996). The *E. coli* ADB box contains sequences between nucleotides 1469-1483 of the 16S rRNA (Sprengart et al., 1996), corresponding to nucleotides 1416-1430 of the tobacco 16S rRNA (Dams et al., 1988; sequence between nucleotides 104173-104187 in Shinozaki et al., 1986).

**Figure 2A.** Base-pairing between plastid ADB and *atpB*, *clpP*, *rbcl*, *psbB* and *psbA* mRNAs (underlined). Multiple alternative DB-ADB interactions are shown. Nucleotides changed to reduce or alter mRNA-rRNA interaction are in lower case. The number of potential nucleotide pairs formed with the 26 nt ADB region is in parenthesis. The number of pairing events affected by mutagenesis is in bold.

**Figure 2B.** Complementarity of Prn T7 phage gene 10 leader derivatives with the *E. coli* and plastid ADB sequences. Nucleotides changed to reduce or alter mRNA-rRNA interaction are in lower case. The number of potential nucleotide pairs formed with the 26 nt ADB region is in parenthesis.

**Figure 3A.** DNA sequence of the chimeric Prn plastid promoter fragments with *atpB* and *clpP* translation control regions. The plasmid name that is the source of the promoter fragment is given in parenthesis. The Prn promoter sequence is underlined; nucleotide at which transcription initiates in tobacco plastids is marked with filled circle; translational

initiation codon (ATG) is in bold; SD is underlined with a wavy line; nucleotides of the 5' and 3' restriction sites and point mutations are in lower case.

5 **Figure 3B.** DNA sequence of the chimeric Prn plastid promoter fragments with *rbcL* and *psbB* translation control regions. For details see description of Fig. 3A.

10 **Figure 3C.** DNA sequence of the chimeric Prn plastid promoter fragments with *psbA* translation control regions. For details see description of Fig. 3A.

15 **Figure 3D.** DNA sequence of the chimeric Prn plastid promoter fragments with the T7 phage gene 10 (PrnLT7g10+DB/Ec) plastid (PrnLT7g10+DB/pt) and synthetic DB (PrnLT7g10-DB). For details see description of Fig. 3A.

20 **Figure 4A.** Plastid transformation vector pPRV111A with chimeric *neo* genes. Plasmid serial numbers, for example pHK34, designate pPRV111A plastid transformation vectors derivatives; adjacent plasmid numbers in parenthesis (e.g. pHK14) designate the source of the chimeric *neo* gene in pUC118 or pBSIIKS+ vectors. Arrows mark orientation of the selectable marker gene (*aadA*) and of the chimeric *neo* gene. Plastid targeting sequences are underlined in bold. Components of the chimeric *neo* genes are: Prn, rRNA operon promoter fragment; L, leader sequence; DB, downstream box; *NheI* site which serves as a synthetic DB is marked by a heavy line; *neo*, neomycin phosphotransferase coding region; 30 *TrbcL*, *rbcL* 3'-untranslated region. *16SrDNA*, *trnV*, *rps12/7* are plastid genes (Shinozaki et al., 1986). The restriction sites marked for: *EcoRI*, *SphI*, *StuI*, *SacI*, *NheI*, *NcoI*, *XbaI*, *HindIII*, *BamHI* and *BglII*. Restriction

sites in brackets were eliminated during construction. The *neo* translation initiation in plasmid pHK36 is included in NcoI site (not marked). The presence and relative order of NheI (\*\*) and NcoI (\*) restriction sites in the plasmid pPRV111A -DB derivatives (pHK35, pHK37, pHK40, pHK42, pHK43) are marked by asterisks. The promoter sequences are shown in Figures 3B, C and D.

**Figure 4B.** Plastid transformation vector pPRV111B with chimeric *neo* genes. See description of Fig. 4A. The promoter sequences are shown in Fig. 3A.

**Figure 5.** Construction of *Prrn* promoter-plastid leader fragments by overlap extension PCR.

**Figure 6.** Construction by the PCR of *Prrn*LT7g10+DB/Ec promoter (SacI-NheI fragment) in plasmid pHK18.

**Figure 7.** Construction by PCR of the *Prrn*LT7g10+DB/pt promoter (SacI-NheI fragment) in plasmid pHK19.

**Figure 8.** Restriction map of plasmids pHK2 and pHK3 with the *Prrn*(L)*rbcL*(S)::*neo*::*TrbcL* gene. Restriction enzyme cleavage sites are marked for: BamHI, EcoRI, HindIII, NcoI, NheI, SacI, XbaI.

**Figure 9.** DNA sequence of the *Prrn*(L)*rbcL*(S)::*neo*::*TrbcL* gene in plasmid pHK3. Plasmid pHK2 carries an identical *neo* gene, except that there is an EcoRI site upstream of the SacI site.

**Figure 10.** NPTII accumulation in tobacco leaves

detected by protein gel blot analysis. Amount of total soluble leaf protein ( $\mu\text{g}$ ) loaded on SDS-PAGE gel is indicated above the lanes. Lanes are designated with plasmid used for plant transformation;  $\mu\text{g}$  protein loaded per lane is given below. NPTII standard and Nt-pTNH32 extracts were run as positive controls; extracts from wild-type non-transformed plants (wt) were used as negative controls.

10        **Figure 11.** The levels of *neo* mRNA in the transplastomic leaves. The blots were probed for *neo* (top) and cytoplasmic 25S rRNA as loading control (bottom). Positions of the monocistronic *neo* mRNA in vector pPRV111A (Figure 4A), the monocistronic *neo* and  
15        dicistronic *neo-aadA* transcripts in vector pPRV111B (Figure 4B) and the monocistronic *neo* and dicistronic *rbcL-neo* transcripts in pTNH32 transformed plants (Carrer et al., 1993) are marked. Lanes are designated with the transgenic plant serial number. 4  $\mu\text{g}$  total  
20        cellular RNA was loaded per lane.

**Figure 12.** Fraction of a codon encoding a particular amino acid and triplet frequency per 1000 codons in the mutagenized *atpB* and *rbcL* DB region.  
25        Altered nucleotides are in lower case.

**Figure 13A.** NPTII accumulation in tobacco roots detected by protein gel blot analysis. Lanes are designated with the plasmid used for plant  
30        transformation;  $\mu\text{g}$  protein loaded per lane is given below. NPTII standard was run as positive control; extracts from wild-type non-transformed plants (wt) were used as negative controls.

Figure 13B. Steady-state levels of *neo* mRNA in tobacco roots. The *neo* probe detects a monocistronic mRNA in plants transformed with vector pPRV111A (Figure 4A), and a monocistronic *neo* and a dicistronic *neo-aadA* transcript in plants transformed with vector pPRV111B (Figure 4B). Lanes are designated with the transgenic plant serial number. 4  $\mu$ g total cellular RNA was loaded per lane.

Figure 14. Protein gel blot analysis to detect NPTII accumulation in tobacco seeds. Lanes are designated with plasmid used for plant transformation;  $\mu$ g protein loaded per lane is given below. NPTII standard was run as positive control; extracts from wild-type non-transformed plants (wt) were used as negative controls.

Figure 15A. Diagram showing integration of the chimeric *neo* and *aadA* genes into the plastid genome by two homologous recombination events via the plastid targeting sequences (underlined). On top is shown a diagram of plasmids pHK30 and pHK32 are plastid transformation vector pPRV111B derivatives (Zoubenko et al., 1994). Horizontal arrows mark gene orientation. For description of chimeric *neo* genes, see Figure 4B. *16SrDNA*, *trnV*, *rps12/7* are plastid genes (Shinozaki et al., 1986). The restriction sites marked for: EcoRI (E), SacI (S), NheI (N), XbaI (X), HindIII (H), BamHI (Ba) and BglII Restriction sites in brackets were eliminated during construction. In the middle the wild-type plastid DNA region (Wt-ptDNA) targeted for insertion is shown. Lines connecting plasmids and ptDNA mark sites of homologous recombination at the end of the vector



plastid-targeting regions. The transformed plastid genome segment (T-ptDNA) map is shown on the bottom.

5       **Figure 15B.** DNA gel blot analysis confirms of integration of the *neo* and *aadA* genes into the plastid genome. The blot on top was probed with the plastid targeting sequence (Probe 1 in Figure 15A). It lights up 4.2-kb and 1.4-kb fragments in transplastomic lines, and a 3.1-kb fragment in wild-type (see Figure 15A). Note that the 1.4-kb signal is weak in most clones. The blot  
10       on the bottom was probed for *neo* sequences, which are present only in the transplastomic lines.

**Figure 16A.** Diagram showing integration of the *bar* gene into the tobacco plastid genome. Map of the plastid targeting region in plasmid pJEK6 is shown on top. The  
15       targeted region of the wild-type plastid genome (wt-ptDNA) is shown in the middle. Integrated transgenes in the transplastome (T-ptDNA) are shown at the bottom. Map positions are shown for: the *bar* gene; *aadA*, the  
20       selectable spectinomycin resistance gene; *16SrDNA* and *rps12/7*, plastid genes (Shinozaki et al., 1986). Arrows indicate direction of transcription. Map position of the probe (2.5 kb) is marked by a heavy line; the wild-type (2.9-kb) and transgenic (3.3-kb, 1.9-kb) fragments  
25       generated by *Sma*I and *Bgl*III digestion are marked by thin lines.

**Figure 16B.** DNA gel blot confirms integration of *bar* into the tobacco plastid genome. Data are shown for transplastomic lines Nt-pJEK6-2A through E, Nt-pJEK6-5A  
30       through E and Nt-pJEK6-13A and B, and the wild-type parental line. *Sma*I-*Bgl*III digested total cellular DNA was probed with the 2.5-kb *Apa*I-*Bam*HI plastid targeting sequence marked with heavy line in Figure 16A.

Figure 17. PAT assay confirms *bar* expression in tobacco plastids. PAT activity was determined by conversion of PPT into acetyl-PPT using radiolabeled  $^{14}\text{C}$ -Acetyl-CoA. Data are shown for transplastomic lines Nt-pJEK6-2D, Nt-pJEK6-5A and Nt-pJEK6-13B, nuclear transformant Nt-pDM307-10 and wild-type (wt).

Figure 18A. Transplastomic tobacco plants are herbicide resistant. Wild-type and pJEK6-transformed plants 13 days after Liberty spraying (5 ml, 2% solution).

Figure 18B. Maternal inheritance of PPT resistance in the seed progeny. Seeds from reciprocal crosses with Nt-pJEK6-5A plants germinated on 0, 10 and 50 mg/L PPT. wt x pJEK6-5A, transplastomic used as pollen parent; pJEK6-5A x wt, transplastomic line female parent. Resistant seedlings are green on PPT medium, sensitive seedlings are bleached.

Figure 19. The engineered bacterial *bar* coding region DNA sequence in plasmid pJEK3 and pJEK6 and encoded amino acid sequence. Nucleotides encoding the *rbcl* five N-terminal amino acids are in lower case. Nucleotides added at the 3' end during construction are also in lower case. NcoI, BglII and XbaI cloning sites are marked.

Figure 20A. The synthetic *bar* gene DNA sequence and the encoded amino acid sequence. The arginines encoded by AGA/AGG codons are in bold. Original nucleotides are in capital letters, altered bases are in lower case. Restriction sites used for cloning are marked.

Figure 20B. The synthetic *s2-bar* gene DNA sequence and the encoded amino acid sequence. The arginines

encoded by AGA/AGG codons are in bold. Original nucleotides are in capital letters, altered bases are in lower case. Restriction sites used for cloning are marked.

5

**Figure 21.** Synthetic and bacterial *bar* genes. The *bar* coding region is expressed in the *Prrn*/*TrbcL* cassettes. Note that the *Prrn* promoters differ with respect to the translational control region.

10

**Figure 22A.** PAT is expressed in *E. coli* from *bar*, but not from *s-bar* coding region. PAT activity was determined by conversion of PPT into acetyl-PPT using radiolabeled <sup>14</sup>C-Acetyl-CoA. Data are shown for *E. coli* transformed with plasmids pJEK6 and pKO12 carrying the *bar* gene, and pKO8, carrying *s-bar*.

15

**Figure 22B.** PAT assay confirms expression of *bar* and *s-bar* in tobacco plastids. PAT activity was determined by conversion of PPT into acetyl-PPT using radiolabeled <sup>14</sup>C-Acetyl-CoA. Data are shown for transplastomic lines Nt-pJEK6-13B and Nt-pKO3-24a,B carrying *bar* and *s-bar*, respectively.

20

**Figure 23A.** Plastid transformation vector with FLARE16-S as selectable marker targeting the plastid inverted repeat region. DNA and protein sequence at the *aadA-gfp* junction. Nucleotides derived from *aadA* and *gfp* are in capital, adapters sequences and the point mutation used to create the *Bst*XI restriction site (bold) are in lower case.

25

30

**Figure 23B.** Physical map of plastid transformation vector with FLARE16-S as selectable marker targeting the plastid inverted repeat region. Shown are: the promoter

(P) and 3'UTR (T) of the *aadA16pt-gfp* coding region and its component parts (*aadA* and *gfp* coding regions); *rrn16* and *rps12/7* plastid genes; restriction endonuclease sites HindIII (removed), SpeI, XbaI, NcoI, BstXI, NheI, EcoRI. In plasmid pMSK56 *aadA16pt-gfp* is expressed from the Pr<sub>rrn</sub>:LatpBDB promoter and encodes FLARE16-S1. In plasmid pMSK57 *aadA16pt-gfp* is expressed from the Pr<sub>rrn</sub>:LrbcLDB promoter and encodes FLARE16-S2.

**Figure 24.** Localization of FLARE16-S to tobacco plastids by laser scanning confocal microscopy in heteroplastomic tissue. Images were processed to detect FLARE16-S (green) and chlorophyll fluorescence (red) and both in a merged view. Sections are shown from plants expressing FLARE16-S1 (a,b) and FLARE16-S2 (3c-f). Note wild-type and transformed plastids in leaves (3a,c,d), chromoplasts of petals (3b), trichomes (3e) and non-green root plastids (f). White arrows mark transplastomic organelles. Bars represent 25  $\mu$ m.

20

**Figure 25.** Immunoblot analysis of FLARE16-S accumulation in chloroplasts. The amount of loaded protein ( $\mu$ g) is indicated above the lanes. Quantification of FLARE16-S1 (Nt-pMSK56 plants) and FLARE16-S2 (Nt-pMSK57 plants) is based on comparison with a purified GFP dilution series. Extract from a wild-type plant (Nt) was used as negative control.

25

**Figure 26A.** Amplification of border fragments confirms integration of FLARE-S genes into the plastid genome. Maps of the plastid targeting regions of the rice (pMSK49) and tobacco (pMSK57) vectors, the segment of the rice and tobacco plastid genomes targeted by the

30

vectors (Os-wt and Nt-wt), and the same regions after integration of FLARE-S genes. The ends of plastid targeting regions are connected with cognate sequences in the wild-type plastid genome. Plastid genes *16SrDNA*, *trnV* and *rps12/7* are marked only in the wild-type plastid genomes. The position of PCR primers (O1-O6) and the PCR fragments generated by them are also shown.

**Figure 26B.** Amplification of border fragments confirms integration of FLARE-S genes into the plastid genome. Gels with PCR-amplified left and right border fragments, and with *aadA* fragment. Results are shown for rice (Os-pMSK49-1 and Os-pMSK49-2) and tobacco (Nt-pMSK57) transplastomic lines and wild-type (Os-wt) rice. The molecular weight markers is EcoRI- and HindIII-digested  $\lambda$  DNA.

**Figure 27.** Localization of FLARE11-S3 to rice chloroplasts in the Os-pMSK49-5 line by laser scanning confocal microscopy. Images were processed to detect FLARE11-S (green) and chlorophyll fluorescence (red) and both in a merged view. Arrows point to mixed populations of plastids in cells. Bar represents 25  $\mu$ m.

**Figure 28.** The sequence of FLARE16-S is shown.

25

**Figure 29.** The sequence of FLARE16-S1 is shown.

**Figure 30.** The sequence of FLARE16-S2 is shown.

**Figure 31.** The sequence of FLARE11-S is shown.

30

**Figure 32.** The sequence of FLARE11-S3 is shown.

Figures 33A and 33B. The sequence of pMSK35 is shown.

Figures 34A and 34B. The sequence of pMSK49 is shown.

Figure 35. A table describing the FLARE constructs of the invention.

#### 10 DETAILED DESCRIPTION OF THE INVENTION

DNA cassettes for high level protein expression in plastids are provided herein. Higher plant plastid mRNAs contain sequences within 50 nt downstream of AUG that are complementary to the 16S rRNA 3-region. These complementary sequences are approximately at the same position as DB sequences in *E. coli* mRNAs. See Figures 1A and 2A. Interestingly, the tentative plastid DB sequence significantly deviates from the *E. coli* DB consensus, since the tobacco plastid and *E. coli* 16S rRNA sequence in the anti-downstream-box (ADB) region is significantly different (Figure 1B). The feasibility of improving protein expression by incorporating DB sequences in plastids was assessed by constructing a series of chimeric 5' regulatory regions consisting of the plastid rRNA operon  $\sigma^{70}$ -type promoter (Prn-114; Svab and Maliga, 1993; Vera and Sugiura, 1995) and the leader sequence of plastid mRNAs with the native DB, mutagenized DB and synthetic DB sequences. The plastid mRNA leaders differ with respect to the presence and position of the SD sequence. Translation efficiency from the chimeric promoters was determined by expressing the bacterial *neo* gene in plastids. The *neo* (or *kan*) gene encodes neomycin phosphotransferase (NPTII) and

confers resistance to kanamycin in bacteria and plastids (Carrer et al., 1993). We have found that NPTII from the chimeric neo transcripts accumulates in the range of 0.2% to 23% of the total soluble leaf protein, indicating the importance of translational control signals in the mRNA 5' region for high-level protein expression.

There is great interest in producing recombinant proteins in plants plastids which, thus far have been expressed from nuclear genes only (Arntzen, 1997; Conrad and Fiedler, 1998; Kusnadi et al., 1997). Protein levels produced from the PrnLrbcL+DBwt and PrnLT7g10 expression cassettes described here significantly exceed protein levels reported for nuclear genes. Accumulation of NPTII from nuclear genes is typically  $\ll 0.1\%$  (Allen et al., 1996), the highest value being 0.4% of the total soluble protein (Houdt et al., 1997). We reported earlier accumulation of 1% NPTII from a plastid neo transgene (Carrer et al., 1993). Other examples for protein accumulation from plastid transgenes are 2.5%  $\beta$ -glucuronidase (GUS) (Staub and Maliga, 1993)) and 3-5% of the *Bacillus thuringiensis* (Bt) crystal toxins (McBride et al., 1995). As compared to this earlier report, we have achieved a significant increase in NPTII levels, up to 23% of total soluble protein.

FLARE-S, a protein obtained by fusing an antibiotic-inactivating enzyme with the *Aequorea victoria* green fluorescence protein accumulated to 8% and 18% of total soluble protein from the PrnLatpB+DBwt and PrnLrbcL+DBwt cassettes provided herein. See Example 8. High-level protein accumulation from the cassettes of the present invention can be clearly attributed to engineering the translational control

region (TCR) of the chimeric genes. These novel genetic elements may be used in different applications to drive expression of proteins with agronomic, industrial or pharmaceutical importance.

5           There is a strong demand for methods that control the flow of transgenes in field crops. Incorporation of the transgenes in the plastid genome rather than the nuclear genome results in natural transgene containment, since plastids are not transmitted via pollen in most  
10 crops (Maliga, 1993). Plastid transformation in crops has not been widely employed due to the lack of technology. Enhanced expression of selective markers should yield higher transformation efficiencies. The chimeric promoters of the present invention facilitate  
15 extension of plastid transformation to agronomically and industrially important crops. Indeed, high-level expression from the PrnLatpB+DBwt cassette described here resulted in ~25-fold increase in the frequency of kanamycin-resistant transplastomic tobacco lines. More  
20 importantly, high levels of marker gene expression following plastid transformation have been obtained in rice, the first cereal species in which plastid transformation has been successful. The results are set forth in Example 8.

25           The following examples are provided to illustrate various embodiments of the present invention. They are not intended to limit the invention in any way.

          The protocols set forth below are provided to facilitate the practice of the present invention.

30



PREPARATION OF CHIMERIC 5' CASSETTES FOR ELEVATED  
EXPRESSION OF HETEROLOGOUS PROTEINS IN PLASTIDS OF  
HIGHER PLANTS

5 Identification of a potential downstream box in  
plastid mRNAs

The presence or absence of downstream box elements  
in mRNA molecules was determined for the following  
genes: *psbB* (Tanaka et al., 1987) and *psbA* (Sugita and  
10 Sugiura, 1984), photosystem II genes; *rbcL*, encoding the  
large subunit of ribulose-1,5-bisphosphate  
carboxylase/oxygenase (Shinozaki and Sugiura, 1982);  
*atpB*, encoding the ATPase  $\beta$  subunit (Orozco et al.,  
1990); and *clpP*, encoding the proteolytic subunit of the  
15 Clp ATP-dependent plastid protease (Hajdukiewicz et  
al., 1997). Interestingly, most or all of the PclpP-53  
promoter is downstream of the transcription initiation  
site, therefore the PrnLclpP constructs are assumed to  
contain two promoters: Prn-114 and PclpP-53.  
20 Transcription initiation sites for these genes were  
described in references cited above; for nucleotide  
position of the genes in the plastid genome see  
Shinozaki et al., 1986.

Initially, it was assumed that the plastid ADB is  
25 similar in size and position as the *E. coli* ADB in the  
16S rRNA. The *E. coli* ADB is localized on a conserved  
stem structure between nucleotides 1469 to 1483 (15 nt)  
that corresponds to nucleotides 1416 and 1430 of the  
plastid 16S rRNA (Dams et al., 1988; Sprengart et al.,  
30 1996). Although in both cases, the ADB is contained in  
the 16S rRNA penultimate stem, the actual ADB sequence  
is different in plastids and in *E. coli* (Figure 1B).  
The N-terminal coding regions of plastid genes *atpB*,  
*clpP*, *rbcL*, *petA*, *psaA*, *psaB*, *psbA*, *psbB*, *psbD* and *psbE* were

searched for potential DB sequences. The homology search was carried out with a 26 nucleotide sequence centered on the tentative DB region (Figure 1B). The search revealed short stretches of imperfect homology with alternative solutions. Since the position of DB in the mRNA is quite flexible (Etchegaray and Inouye, 1999), we show four potential DB-ADB interactions for *atpB* and *rbcL* in Figure 2A. Two plastid mRNAs were selected to test the role of DB in the translation of plastid mRNAs:

1) *atpB* mRNA lacks a SD sequence; and 2) *rbcL* mRNA contains a SD sequence at the prokaryotic consensus. In addition, the phage T7 gene 10 (T7g10) leader was included in the study. This leader has a well-characterized *E. coli* DB sequence (Figure 2B; Sprengart et al., 1996). Additional plastid mRNAs with potential DB sequences shown in Figure 2A are *clpP*, *psbB* and *psbA*.

#### Experimental strategy to test the efficiency of leader sequences for translation

To compare the efficiency of translation from the 5'-UTR of the selected genes, the 5'-UTR was cloned downstream of the strong plastid rRNA operon  $\sigma^{70}$ -type promoter (Prn-114) (Svab and Maliga, 1993; Allison et al., 1996), which initiates transcription from multiple adjacent nucleotides (-114, -113, -111; Sriraman et al., 1998). The promoter fragments were constructed as SacI-NheI or a SacI-NcoI fragments. Construction of the chimeric promoters using conventional molecular biological techniques is set forth in detail in the next section.

Two constructs were prepared for each 5'-UTR selected: one with (+DB) and one without (-DB) a native

downstream box. It will be obvious from the forthcoming discussion, that the -DB constructs have a synthetic DB provided by the NheI restriction site. The promoters were cloned upstream of the coding region of a kanamycin resistance (*neo*) gene, which is available on an NheI-XbaI or NcoI-XbaI fragment. For the stabilization of the mRNA, the *rbcL* gene 3'-untranslated region was cloned downstream of *neo* as an XbaI-HindIII fragment. The chimeric *neo* genes can therefore be excised from the pUC118 or pBSIIKS+ plasmids as SacI-HindIII fragments. These source plasmids are listed in Table 1.

Table 1. Salient features of chimeric promoters<sup>a</sup>.

	Source of 5'-UTR (nucleotides from AUG)	SD	DB	Promoter fragment	pUC118 (U) or pBSIIKS <sup>+</sup> (B)	pPRV111A, B
15	atpB (-90/+42)	-	wt	SacI/NheI	pHK10 (U)	pHK30 (B)
	atpB (-90/+6)	-	s	SacI/NheI	pHK11 (U)	pHK31 (B)
20	atpB (-90/42)	-	m	SacI/NheI	pHK50 (B)	pHK60 (B)
	clpP (-53/+48)	-	wt	SacI/NheI	pHK12 (U)	pHK32 (B)
	clpP (-53/+6)	-	s	SacI/NheI	pHK13 (U)	pHK33 (B)
25	rbcL (-58/+42)	+	wt	SacI/NheI	pHK14 (B)	pHK34 (A)
	rbcL (-58/+6)	+	s	SacI/NheI	pHK15 (U)	pHK35 (A)
	rbcL (-58/+42)	+	m	SacI/NheI	pHK54 (B)	pHK64 (A)
30	psbB (-54/+45)	+	wt	SacI/NheI <sup>d</sup>	pHK16 (U)	pHK36 (A)
	psbB (-54/+3)	+	s	SacI/NcoI <sup>d</sup>	pHK17 (U)	pHK37 (A)
	<sup>b</sup> T7g10+DB/Ec (-63/+24)+		Ec	SacI/NheI	pHK18 (B)	pHK38 (A)
	<sup>b</sup> T7g10+DB/pt (-63/+24)+		pt	SacI/NheI	pHK19 (B)	pHK39 (A)
35	T7g10-DB (-63/+9) +		s	SacI/NheI	pHK20 (B)	pHK40 (A)
	psbA (-85/+21)	-	wt	SacI/NheI	pHK21 (U)	pHK41 (A)
	psbA (-85/+3)	-	s	SacI/NcoI <sup>e</sup>	pHK22 (U)	pHK42 (A)
	<sup>c</sup> psbA(+GC) (-85/+3)	-	-	sSacI/NcoI <sup>e</sup>	pHK23 (U)	pHK43 (A)

<sup>a</sup>-SD+, SD at prokaryotic consensus position; SD-, no SD at prokaryotic consensus position;

DB wt, wild-type; m, mutants; s, NheI site as synthetic DB.

<sup>b</sup>Ec or pt refers to construct with *E. coli* or plastid DB sequence.

<sup>c</sup>psbA(+GC) indicates addition of GC to the wild-type A at the mRNA 5'-end.

<sup>d</sup>In source gene *psbB* translation initiation codon is within NcoI site; therefor +DB construct pHK16 has this NcoI site upstream of the NheI site; see Figure 9.

<sup>e</sup>Translation initiation codon is included in NcoI site; NheI site is directly downstream in *kan* coding region; see Figure 8.

The *Prrn* promoter fragment is available in plasmid pPRV100A (Zoubenko et al., 1994). The promoters were designed to include sequences between -197 nt and -114 nt upstream of the mature 16S rRNA 5' end. Nucleotide -197 is the 5'-end of the *Prrn* promoter constructs utilized for these and other studies (Svab and Maliga, 1993; -1 is the first nucleotide upstream of the mature 16S rRNA). The G at the -114 position is one of three transcription initiation sites; the other two are the adjacent C (-113) and A (-111) nucleotides (Allison et al., 1996, Sriraman et al., 1998). The nucleotide at which *Prrn* transcription would initiate is marked by a filled circle in Figure 3A-D. In most constructs, this is a G (-114) as in the native promoter. In two constructs the G was replaced by an A, as in the *psbA* promoter which is the source of the leader sequence (pHK21, pHK22; see below).

## 20 DESIGN OF THE 5' LEADER FROM *atpB*

For the *atpB* gene, multiple mRNA 5'-ends were mapped in tobacco leaves including at least four primary transcripts indicating transcription from four promoters and a processed 5'-end 90 nucleotides upstream of the translation initiation codon (Orozco et al., 1990). The terminal nucleotide of the processed *atpB* 5'-end is a G. Therefore, the chimeric *PrrnLatpB* promoters were designed to initiate transcription at a G, anticipating that the leader sequence of the chimeric transcript will be a perfect reproduction of the processed *atpB* mRNA 5'-end. Out of the *atpB* coding region, 42 and 6 nucleotides are included in the +DBwt and -DB constructs, respectively. The 42 nucleotides include

four potential DB sequences shown in Figure 2A. Two point mutations in the leader sequence were designed to eliminate NheI (T to A) and EcoRI (G to A) restriction sites without affecting the predicted mRNA 5' secondary structure. In the -DB constructs, two codons (6 nucleotides) were retained from the native coding region upstream of the NheI restriction site (GCTAGC sequence) in which the stop codon is out-of-frame (Figure 3A). Eleven silent point mutations were introduced in the DB region of the PrnLatpB+DBm construct to either minimize the number of base pairs, or to change the nature of base pairing (for example G-C to G-U) (Figure 2A; Figure 3A).

#### 15 DESIGN OF THE 5' LEADER FROM *clpP*

Two major mRNA 5'-ends of the *clpP* gene were mapped in tobacco leaves (Hajdukiewicz et al., 1997). The terminal nucleotide of the proximal primary transcript is a G. Therefore, the chimeric PrnLclpP promoters were designed to initiate transcription at a G, anticipating that the leader sequence of the chimeric transcript will be a perfect reproduction of the leader transcribed from the Pclp-53 NEP promoter. Out of the *clpP* coding region, 48 and 6 nucleotides are retained in the +DBwt and -DB constructs, respectively. The 48 nucleotides include four potential DB sequences as shown in Figure 2A. In the -DB constructs, two codons (6 nucleotides) were retained from the native coding region upstream of the NheI restriction site (GCTAGC sequence) in which the stop codon is out-of-frame.

**DESIGN OF THE 5' LEADER FROM *rbcL***

One primary and one processed mRNA 5'-end were mapped in tobacco leaves for the *rbcL* gene (Shinozaki and Sugiura, 1982). The terminal nucleotide of the processed 5' end is a T. The chimeric PrnLrbcL promoters were designed to initiate transcription at a G, one nucleotide downstream of the terminal T. Forty-two and 6 nucleotides out of the *rbcL* coding region are included in the +DB and -DB constructs, respectively. The 42 nucleotides include four potential DB sequences as shown in Figure 2A. The one point mutation (G to A) in the leader sequence was designed to eliminate an EcoRI restriction site without affecting the predicted mRNA 5' secondary structure. In the -DB constructs, two codons (6 nucleotides) were retained from the native coding region upstream of the NheI restriction site (GCTAGC sequence) in which the stop codon is out-of-frame. Twelve silent point mutations were introduced into the DB region of the PrnLrbcL+DBm construct to either minimize the number of base pairs, or to change the nature of base pairing (for example G-C to G-U) (Figure 2A, Figure 3B).

**DESIGN OF THE 5' LEADER FROM *psbB***

One primary and one processed mRNA 5'-end for the *psbB* gene were tentatively identified in tobacco leaves (Tanaka et al., 1987). The leader sequence was designed to initiate transcription from the G (-114) of the Prn promoter, and include the intact secondary (stem) structure assumed to be involved in stabilizing the mRNA. Forty-five and 3 nucleotides out of the *psbB* coding region are included in the +DB and -DB constructs, respectively. The 45 nucleotides include

four potential DB sequences shown in Figure 2A. Since the ATG is naturally included in an NcoI site that is used to fuse the *neo* coding region with the *psbB* leader, no amino acid from the *psbB* coding region is added in the -DB construct.

#### DESIGN OF THE 5' LEADER FROM *psbA*

One mRNA 5'-end was mapped for the *psbA* gene in tobacco leaves (Sugita and Sugiura, 1984). The terminal nucleotide of the primary transcript is an A. Therefore, the chimeric PrrnLpsbA promoters were designed to initiate transcription at an A, anticipating that the leader sequence of the chimeric transcript will be a perfect reproduction of the leader transcribed from the *psbA* promoter. Twenty-one and 3 nucleotides out of the *psbA* coding region are included in the +DB and -DB constructs, respectively. The 21 nucleotides include the potential DB sequence as shown in Figure 2A. Since the *neo* coding region was linked to the chimeric promoter via an NcoI site which includes the translation initiation codon (ATG), no amino acid from the *psbA* coding region is added in the -DB constructs. This is true of a second -DB promoter, in plasmid PHK23, in which transcription is designed to initiate from the Prrn G (-114) and C (-113) (Figure 3C).

#### DESIGN OF THE T7 PHAGE GENE 10 LEADER

The T7 phage gene 10 leader (63 nucleotides) was shown to promote efficient translation initiation in *E. coli* (Olins et al., 1988). This leader is used in the *E. coli* pET expression vectors (Studier et al., 1990; Novagen Inc.). The terminal nucleotide at the 5'-end is a G. Therefore, the chimeric PrrnT7g10L promoters were

designed to initiate transcription at a G, anticipating that the leader sequence of the chimeric transcript will be a reproduction of the T7 phage gene 10 mRNA, with the exception of a T to A mutation which was introduced to eliminate an XbaI site. Twenty-four and 9 nucleotides from the T7 phage gene 10 coding region are included in the +DB/Ec (with *E. coli* DB sequence) and -DB constructs, respectively. To compare the efficiency of *E. coli* and plastid DB sequences in plastids, a second +DB promoter was constructed with the tobacco DB sequence (PrrnT7g10L+DB/pt). The native T7g10 leader has an NheI site directly downstream of the translation initiation codon. This NheI site was removed by a T to A point mutation in the +DB constructs (Figure 3D).

For introduction into the plastid genome, the chimeric *neo* genes were cloned into plastid transformation vector pPRV111A or pPRV111B. See U.S. Patent 5,877,402, the disclosure of which is incorporated herein by reference. The pPRV111 vectors target insertions into the inverted repeat region of the tobacco plastid genome, and carry a selectable spectinomycin (*aadA*) resistance gene. The sequences of the vectors have been deposited in GenBank (U12812, U12813). The chimeric *neo* gene in vector pPRV111B is in tandem with the *aadA* gene, whereas in vector pPRV111A the chimeric *neo* is oriented divergently. The general outline of the plastid transformation vector with the chimeric *neo* genes is shown in Figures 4A and 4B.



# CONSTRUCTION OF CHIMERIC Prn<sup>+</sup> PROMOTERS WITH PLASTID MRNA LEADERS

The chimeric Prn<sup>+</sup> promoter/leader fragments were constructed as a SacI-NheI or SacI-NcoI fragments (Table 1, below) by overlap extension PCR (SOE-PCR), essentially as described in Lefebvre et al., (1995). Construction of the Prn<sup>+</sup>-plastid leader segments is schematically shown in Figure 5. The objective of the PCR-1 step is to 1) amplify the Prn<sup>+</sup> promoter fragment while 2) adding a SacI site upstream and a seam-less overlap with the specific downstream leader sequence. The reaction contains: 1) a primer (oligonucleotide) to add a SacI site at the 5'-end of the fragment; 2) a suitable template containing the Prn<sup>+</sup> promoter sequence in plasmid pPRV100A (Zoubenko et al., 1994); and 3) a primer to add on the overlap with the leader sequence at the 3' of the amplified product. The objective of the PCR-2 step is to create the chimeric promoter with DB sequence using: 1) the product of PCR-1 step as a primer; 2) a suitable DNA template containing the specific leader sequence; and 3) primer (oligonucleotide) to include NheI restriction site at the 3'-end of the amplification product. The product of the PCR-2 is the SacI-NheI chimeric Prn<sup>+</sup> promoter fragment with DB sequence. The objective of the PCR-3 step is to remove the DB sequence while introducing a suitable NheI or NcoI restriction site. The product of PCR-3 is the SacI-NheI or SacI-NcoI chimeric Prn<sup>+</sup> promoter fragment in which the DB sequence is replaced with the NheI site. The objective of the PCR-4 step is to replace the wild-type DB with a mutant DB. The product of PCR-4 is a SacI-NheI Prn<sup>+</sup> promoter fragment.

The primers (oligonucleotides) used for the

construction of chimeric promoters are listed in Table 2. The chimeric promoters were obtained by overlap extension PCR using oligonucleotides and DNA templates schematically shown in Figure 5.

5

Table 2.

Oligonucleotides used for the construction of chimeric promoters.

10	#1: 5'-CCCGAGCTCGCTCCCCGCCGTCGTTC-3'
	#2: 5'- CGAATTTAAATAAATGTCCGCTTGACGTCGATCGGTTAATTCTCCAGAAATATAGCCATCC-3'
15	#3: 5'-CCCGCTAGCCGTGGAAACCCAGAAC-3'
	#4: 5'-CCCGCTAGCTCTCATAATAATAAAATAAATAATATGTC-3'
	#5: 5'-TCACTTTGAGGTGGAAACGTAACCTCCAGAAATATAGCCATCC-3'
20	#6: 5'-CCCGCTAGCTTCCTCTCCAGGACTTCG-3'
	#7: 5'-CCCGCTAGCAGGCATTAAATGAAAGAAAGAAC-3'
25	#8: 5'-TAAGAATTTTCACAACAACAAGGTCTACTCGACTCCAGAAATATAGCCATCC-3'
	#9: 5'-CCCGCTAGCTTTGAATCCAACACTTGCTTTAG-3'
	#10: 5'-CCCGCTAGCTGACATAAATCCCTCCCTAC-3'
30	#11: 5'-CAAAGATAAATAGACACTACGTAACCTTATTGCATTGCTCCAGAAATATAGCCATCC-3'
	#12: 5'-CCCGCTAGCATCATTCAATACAACGGTATGAACACG-3'
35	#13: 5'-TTCTAGTGGGAAACCGTTGTGGTCTCCCTCCAGAAATATAGCCATCC-3'
	#14: 5'-CCCGCTAGCCATATGTATATCTCCTTCTTAAAG-3'
40	#15: 5'-CCCGCTAGCCTGTCCACCAGTCATGCTTGCCATA-3'
	#16: 5'-CCCGCTAGCCAAGGCAGGGCTAGTGATTGCCATATGTATATCTCCTTC-3'
	#17: 5'-TTTGTTTAACTTTAAGAAGGAGATATACATATGGCAAGCATGACTGGTGG-3'
45	#18: 5'-CTCCTTCTTAAAGTTAAACAAAATTATTTCTAGTGGGAAACCGTTGT-3'
	#19: 5'-CAAAATAGAAAATGGAAGGCTTTTGTCTCCAGAAATATAGCCATCCC-3'
50	#20: 5'-CAAAATAGAAAATGGAAGGCTTTTTTCCAGAAATATAGCCATCCC-3'
	#21: 5'-GGGCCATGGTAAAATCTTGGTTTATTTAATC-3'

#22: 5'-GGGGCTAGCTCTCTCTAAAAATTGCAGT-3'

#23: 5'-GAATAGCCTCTCCACCCA-3'

5 #24: 5'-CCCGCTAGCCGTGGACACCCCACTTCCACTTGTGTGCGGGTTTATTCTCAT-3'

#25: 5'-CCCGCTAGCTTTGAATCCTACTGAGGCTTTGTTTCTGTTTGAGGACTCAT-3'

10

#### CONSTRUCTION OF CHIMERIC Prnn PROMOTER/atpB LEADER SEGMENTS

PrnnLatpB+DBwt in plasmid pHK10 (Product of PCR-2)

PrnnLatpB-DB in plasmid pHK11 (Product of PCR-3)

15 PrnnLatpB+DBm in plasmid pHK50 (Product of PCR-4)

PCR-1: Oligonucleotides #1, #2 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

PCR-2: Product of PCR-1 step, Oligonucleotide #3 as primers; plasmid pIK79 (see below) DNA as template.

20 PCR-3: Oligonucleotide #1, #4 as primers; Product of PCR-2 step as template.

PCR-4: Oligonucleotide #1, #24 as primers; Product of PCR-2 step as template.

25 Plasmid pIK79 is a Bluescript BS+ phagemid derivative which carries a PvuII/XhoI tobacco plastid DNA fragment between nucleotides 55147-60484 containing the rbcL-atpB intergenic region with divergent promoters for these genes (Shinozaki et al., 1986).

#### 30 CONSTRUCTION OF CHIMERIC Prnn PROMOTER/clpP LEADER SEGMENTS

PrnnLclpP+DBwt in plasmid pHK12 (Product of PCR-2)

PrnnLclpP-DB in plasmid pHK13 (Product of PCR-3)

35 PCR-1: Oligonucleotides #1, #5 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

PCR-2: Product of PCR-1 step, Oligo #6 as primers; tobacco Sal8 ptDNA fragment (Shinozaki et al., 1986) as

template.

PCR-3: Oligonucleotide #1, #7 as primers; Product of PCR-2 step as template.

5

**CONSTRUCTION OF CHIMERIC Prnn PROMOTER/rbcL LEADER SEGMENTS**

PrnnLrbcL+DBwt in plasmid pHK14 (Product of PCR-2)

PrnnLrbcL-DB in plasmid pHK15 (Product of PCR-3)

10 PrnnLrbcL+DBm in plasmid pHK54 (Product of PCR-4)

PCR-1: Oligonucleotides #1, #8 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

15 PCR-2: Product of PCR-1 step, Oligonucleotide #9 as primers; plasmid pIK79 DNA (see description of pHK10 above) as template.

PCR-3: Oligonucleotide #1, #10 as primers; Product of PCR-2 step as template.

PCR-4: Oligonucleotide #1, #25 as primers; Product of PCR-2 step as template.

20

**CONSTRUCTION OF CHIMERIC Prnn PROMOTER/psbB LEADER SEGMENTS**

PrnnLpsbB+DBwt in plasmid pHK16 (Product of PCR-2)

25 PrnnLpsbB-DB in plasmid pHK17 (Promoter from pHK16, digested with SacI/NcoI)

PCR-1: Oligonucleotides #1, #11 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

30 PCR-2: Product of PCR-1 step, Oligo #12 as primers; tobacco Sal8 ptDNA fragment (Shinozaki et al., 1986) as template.

PCR-3 was not necessary, since the psbB translation initiation codon is naturally included in an NcoI site. Therefore, the -DB derivative could be obtained by

SacI/NcoI digestion of the PCR-2 step.

#### CONSTRUCTION OF CHIMERIC Prnn PROMOTER/psbA LEADER SEGMENTS

- 5 PrnnLpsbA+DBwt in plasmid pHK21 (Product of PCR-2)  
PrnnLpsbA -DB in plasmid pHK22 (Product of PCR-3)  
PCR-1: Oligonucleotides #1, #20 as primers; plasmid  
pPRV100A (Zoubenko et al., 1994) DNA as template.  
PCR-2: Product of PCR-1 step, Oligo #22 as primers;  
10 tobacco Sal3 ptDNA fragment (Shinozaki et al., 1986) as  
template.  
PCR-3: Oligonucleotide #1, #21 as primers; Product of  
PCR-2 step as template.
- 15 PrnnLpsbA(GC) -DB in plasmid pHK23 (Product of PCR-2)  
PCR-1: Oligonucleotides #1, #19 as primers; plasmid  
pPRV100A (Zoubenko et al., 1994) DNA as template.  
PCR-2: Product of PCR-1 step, Oligo #21 as primers;  
tobacco Sal3 ptDNA fragment (Shinozaki et al., 1986) as  
20 template.

- In all of the above, PCR amplification was carried out with AmpliTaq DNA polymerase (Perkin Elmer) or Pfu DNA polymerase (Stratagene) and "stepdown" PCR that
- 25 utilizes gradually decreasing annealing temperatures was performed (Hecker and Roux, 1996). The exact amplification conditions for the chimeric Prnn::LatpB promoters are given below. The amplification conditions for the remaining chimeric Prnn - plastid leader  
30 promoters were calculated according to Hecker and Roux (1996), and differ only in the annealing temperatures. Description of PCR conditions for the construction of the chimeric Prnn promoters with plastid mRNA leaders is

given below; for interpretation of individual steps see scheme in Figure 5.

5

PCR-1 Program: 50 picomoles of both primers per 100  $\mu$ l

	1.1 Denature	5 min. at 94 °C	
	2.1 Denature	1 min. at 94 °C	
	2.2 Annealing	0.5 min. at 72 °C	3 cycles
10	2.3 Extension	0.5 min. at 72 °C	
	3.1 Denature	1 min. at 94 °C	
	3.2 Annealing	0.5 min. at 69 °C	3 cycles
	3.3 Extension	0.5 min. at 72 °C	
	4.1 Denature	1 min. at 94 °C	
15	4.2 Annealing	0.5 min. at 66 °C	3 cycles
	4.3 Extension	0.5 min. at 72 °C	
	5.1 Denature	1 min. at 94 °C	
	5.2 Annealing	0.5 min. at 63 °C	3 cycles
	5.3 Extension	0.5 min. at 72 °C	
20	6.1 Denature	1 min. at 94 °C	
	6.2 Annealing	0.5 min. at 60 °C	3 cycles
	6.3 Extension	0.5 min. at 72 °C	
	7.1 Denature	1 min. at 94 °C	
	7.2 Annealing	0.5 min. at 57 °C	20 cycles
25	7.3 Extension	0.5 min. at 72 °C	
	8.1 Extension	10 min. at 72 °C	
	8.2	1 min. at 30 °C	

The PCR-2 program was essentially identical to the PCR1  
 30 program set forth above with the following  
 modifications: 1) Primers in 100  $\mu$ l were the products of  
 1st PCR reaction, 50 picomoles of the oligonucleotide  
 primer were used; and 2) the annealing temperature in

stepdown PCR was from 67 °C to 52 °C. Accordingly, the following annealing temperatures were used: Step 2.2, 67 °C; Step 3.2, 64 °C; Step 4.2, 61 °C; Step 5.2, 58 °C; Step 6.2, 55 °C; Step 7.2, 52 °C.

5

The PCR-3 and PCR-4 programs were essentially identical to the PCR1 program with the following modification:

1) The annealing temperature in stepdown PCR was from 69 °C to 44 °C. Accordingly, the following annealing  
10 temperatures were used: Step 2.2, 69 °C; Step 3.2, 64 °C; Step 4.2, 59 °C; Step 5.2, 54 °C; Step 6.2, 49 °C; Step 7.2, 44 °C. In cases where the yield of the final PCR reaction was too low for efficient cloning, final product was amplified using primers which were used to  
15 generate the ends. The final PCR products were digested with the appropriate restriction enzymes (SacI and NheI or SacI and NcoI) and cloned in plasmids pHK2 or pHK3 (see below).

## 20 CONSTRUCTION OF CHIMERIC PROMOTERS WITH T7 PHAGE GENE 10 mRNA LEADER SEGMENT

The chimeric Prrn promoter/T7gene10 leader (PrrnLT7g10) fragments were constructed as SacI-NheI fragments (Table 1, below).

25

### PrrnLT7g10+DB/Ec promoter in plasmid pHK18

In the absence of a proper DNA template, the PrrnLT7g10+DB/Ec was constructed by employing a modified polymerase chain reaction (Uchida, 1992) in two PCR  
30 steps, as schematically shown in Figure 6. The PCR-1A and PCR1B steps generate two fragments in two separate reactions (A and B). The objective of the PCR-1A step is to amplify Prrn promoter fragment while: 1) adding a

SacI site upstream (Oligonucleotide #1 in Table 2); and  
2) a seam-less overlap with the specific downstream  
leader sequence (Oligonucleotide #13 in Table 2) using  
plasmid pPRV100A (Zoubenko et al., 1994) as DNA  
5 template. The objective of the PCR-1B step is to  
amplify part of the T7g10 leader sequence using  
overlapping oligonucleotides #15 and #17 in Table 2. The  
NheI site is introduced in oligonucleotide #15. Both  
PCR-1A and PCR-1B reactions were carried out by stepdown  
10 PCR as described above for the construction of the  
chimeric Prn promoters.

PCR-2 reaction generating this chimeric promoter  
contained:

- 15 a) The products of the PCR-1A and PCR-1B reactions as  
DNA templates;
- b) Oligonucleotide #18 (0.5 picomole; Table 2) to  
generate overlapping fragments with products of the PCR-  
1A and PCR-1B reactions;
- 20 c) Oligonucleotides #1 and #15 (Table 2) for  
amplification of the final product, 50 picomoles each in  
100  $\mu$ l final volume.

Promoter was amplified by stepdown PCR, as  
described for the chimeric Prn promoters above; the  
25 annealing temperatures were between 72 °C to 57 °C.

PrnLT7g10+DB/pt promoter in plasmid pHK19

The promoter fragment was obtained in one PCR step as  
shown in Figure 7. The reaction contained:

- 30 a) The product of the PCR-2 reaction generating promoter  
PrnLT7g10+DB/Ec in plasmid pHK18 as DNA template; and  
b) Oligonucleotides #1 and #16 (Table 2), 50 picomoles  
each in 100  $\mu$ l final volume.



Promoter was amplified by stepdown PCR, as described for the construction of chimeric Prn promoters above; the annealing temperatures were between 72 °C to 52 °C.

5

PrnLT7g10-DB promoter in plasmid pHK20

The promoter fragment was obtained in one PCR step, which is similar to the PCR-3 step in Figure 5. The reaction contained:

- 10 a) The product of the PCR-2 reaction generating promoter PrnLT7g10+DB/Ec in plasmid pHK18 as DNA template; and  
b) Oligonucleotides #1 and #14 (Table 2), 50 picomoles each in 100 µl final volume.

15 Promoter was amplified by stepdown PCR, as described for the chimeric Prn promoters above; the annealing temperatures were between 72 °C to 52 °C.

The final PCR products were digested with the SacI and NheI restriction enzymes and cloned in plasmid pHK3  
20 to obtain plasmids pHK18, pHK19, pHK20.

**Construction of chimeric neo genes**

Construction of the chimeric promoters was described in the preceding sections. For determining  
25 effects on levels of protein accumulation, the promoters were cloned upstream of a kanamycin-resistance encoding construct, consisting of the neo coding region and the 3'-UTR of the plastid rbcL gene. Such constructs are available in plasmids pHK2 and pHK3, which carry the  
30 same Prn(L)rbcL(S)::neo::TrbcL gene as a SacI-HindIII fragment. Plasmid pHK2 is a pUC118 vector derivative; pHK3 is a pBSIIKS+ derivative. Plasmid maps with relevant restriction sites are shown in Figure 8. DNA

sequence of the *neo* gene in plasmids pHK2 and pHK3 is shown in Figure 9. Note, that in plasmid pHK2 the *neo* gene has an EcoRI site upstream of the SacI site (Figure 8). Prrn and TrbcL have been described by Staub and  
5 Maliga, 1994; the *neo* gene derives from plasmid pSC1 (Chaudhuri and Maliga, 1996). The pUC118 and pBSIIKS+ plasmid derivatives which carry the various promoter constructs are listed in Table 1.

To determine the DNA sequence of the promoter  
10 fragments, the plasmids were purified with the QIAGEN Plasmid Purification Kit following the manufacturer's recommendations. DNA sequencing was carried out using a T7 DNA sequencing kit (version 2.0 DNA , Amersham Cat. No. US70770) and primer No. #23 in Table 2, which is  
15 complementary to the *neo* coding sequence. These promoter sequences are shown in Figure 3A-D.

#### **Introduction of chimeric *neo* genes into the tobacco plastid genome**

20 Suitable vectors are available for the introduction of foreign genes into the tobacco plastid genome. Such vectors are pPRV111A and pPRV111B, which carry a selectable spectinomycin-resistance (*aadA*) gene and target insertions into the repeated region of the  
25 plastid genome (Zoubenko et al., 1994). The chimeric *neo* genes were cloned into one of these plastid transformation vectors (Table 1) and introduced into the tobacco plastid genome by the biolistic process. From the transformed cells plants were regenerated by  
30 standard protocols (Svab and Maliga, 1993). A uniform population of transformed plastid genome copies was confirmed by Southern analysis.

For Southern analysis, total cellular DNA was

prepared by the CTAB method (Saghai-Maroo et al., 1984). Two leaves of each transformed plant were homogenized and incubated at 60°C for 30 minutes in a buffer containing 2% CTAB (tetradecyl-trimethyl-ammonium bromide), 1.4 M NaCl, 20 mM EDTA (pH 8.0), 1 mM Tris/HCl (pH 8.0) and 100 mM  $\beta$ -mercaptoethanol. After chloroform extraction, the DNA was precipitated with isopropyl alcohol and dissolved in water or in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). DNA digested with an appropriate restriction enzyme was electrophoresed on 0.8% agarose gel and transferred to nylon membrane using PosiBlot Transfer apparatus (Stratagene). The blots were probed using Rapid Hybridization Buffer and plastid targeting sequences as a probe labeled with random priming ( $^{32}$ P, Boehringer Mannheim Cat No. 1004760).

Plastid transformation was achieved with each of the plasmids listed in Table 1. Exceptions were plasmids pHK41 and pHK42. It appears that NPTII expression with the psbA leader derivatives was so high that the plants were not viable. It follows that these same leaders may be used to advantage when fused with weaker promoters.

Transplastomic lines are designated by Nt (N. tabacum, the species), the plasmid name (for example pHK30) and an individual line number and a letter identifying regenerated plants. For example, the Nt-pHK30-1D and Nt-pHK30-1C plants were both obtained by transformation with plasmid pHK30, are derived from the same transformation event and were regenerated from the same culture. Nt-pHK30-2 plants are derived from an independent transformation event. Normally, several transformed lines per construct were obtained. However, data are shown here only for one: Nt-pHK30-1D, Nt-pHK31-1C, Nt-pHK60-5A, Nt-pHK32-2F, Nt-pHK33-2A, Nt-pHK34-9C,

Nt-pHK35-4A, Nt-pHK64-3A, Nt-pHK36-1C, Nt-pHK37-2D, Nt-pHK38-2E, Nt-pHK39-3B, Nt-pHK40-12B and Nt-pHK43-1C.

5     **Testing mRNA accumulation by RNA gel blot (Northern)  
analysis**

RNA gel blot analysis was performed to determine steady-state levels of chimeric mRNA in the transplastomic lines. Total leaf RNA was prepared from  
10 the leaves and roots of plants grown in sterile culture according to Stiekema et al (1988). RNA (4 µg per lane) was electrophoresed on 1% agarose gel and transferred to nylon membranes using the PosiBlot Transfer apparatus (Stratagene). The blots were probed using Rapid  
15 Hybridization Buffer Amersham) with a <sup>32</sup>P-labeled neo probe (Pharmacia, Ready-To-Go Random Priming Kit). The neo probe was obtained by isolating the NheI/XbaI fragment from plasmid pHK2. The template for probing the tobacco cytoplasmic 25S rRNA was a fragment which  
20 was PCR amplified from total tobacco cellular DNA with primers 5'-TCACCTGCCGAATCAACTAGC-3' and 5'-GACTTCCCTTGCCTACATTG-3'. RNA hybridization signals were quantified using a Molecular Dynamics PhosphorImager, and normalized to the 25S rRNA signal.

25

**Testing NPTII accumulation by protein gel blot (Western)  
analysis**

Total soluble protein was extracted from the leaves, roots or seeds of transgenic tobacco plants  
30 grown in sterile culture. In case of leaves grown in sterile culture, about 200 mg leaf tissue was homogenized in 1 ml of buffer containing 50 mM Hepes/KOH (pH 7.5), 1 mM EDTA, 10 mM potassium acetate, 5 mM

magnesium acetate, 1 mM dithiothreitol and 2 mM PMSF. The homogenate was centrifuged twice at 4 °C to remove insoluble material. Protein concentration was determined using the Biorad Protein Assay reagent kit. Transgenic tobacco plants expressing *neo* in the plastid genome (Nt-pTNH32-70, Carrer et al., 1993) and wild type plants were used as positive and negative controls, respectively. Proteins were separated in SDS polyacrylamide gels (SDS-PAGE; 15% acrylamide, 6 M urea) and transferred to nitrocellulose membranes using a semi-dry transfer apparatus (Bio-Rad). After blocking non-specific binding sites, the membrane was incubated with 4,000-fold diluted polyclonal rabbit antiserum raised against NPTII (5Prime-3Prime Inc.). HRP-conjugated secondary antibody, diluted 20,000 fold, and ECL chemiluminescence were used for immunoblot detection on X-ray film. NPTII was quantified on the immunoblots by comparison of the experimental samples with a dilution series of commercial NPTII (5Prime-3Prime).

20

#### EXAMPLE 1

**DB sequences enhance protein accumulation from *rbcL* leader; protein accumulation from the *atpB* translation control signals is high but DB-independent**

The role of DB sequences in mRNA translation was tested using *neo* as the reporter gene. The *neo* gene encodes the bacterial enzyme neomycin phosphotransferase (NPTII) (Beck et al., 1982). The tested *neo* genes have the same promoter (Prn) and transcription terminator (TrbcL), and differ only with respect to the translation control region (TCR) comprising the 5' untranslated region of the mRNA and the coding region N-terminus. Two constructs were prepared with the *atpB* and *rbcL* TCRs.

30

One construct contained the wild-type TCR, including the processed 5' untranslated region and 42 nucleotides of the coding region N-terminus (PrrnLatpB+DBwt, plasmid pHK30, Figure 4B; PrrnLrbcL+DBwt, plasmid pHK34, Figure 4A). The second construct contained silent mutations in the 42-nucleotide segment of the *atpB* and *rbcL* N-terminal coding regions to either eliminate or alter mRNA and rRNA base pairing (PrrnLatpB+DBm plasmids pHK60, Figure 2A and Figure 4B; PrrnLrbcL+DBm, pHK64, Figure 2A and Figure. 4A). The silent mutations altered the mRNA sequence without effecting the amino acid sequence. For example, 13 potential base pairs may form between the wild-type *atpB* mRNA and the ADB sequence shown at the bottom in Figure 2A. The 11 silent mutations affect eight base-pairing events for this particular ADB-DB interaction. After mutagenesis, there is a possibility for ten base pairing events, most of which are new. The chimeric *neo* genes were introduced into the tobacco plastid genome by homologous targeting using the biolistic approach (Svab and Maliga, 1993; Zoubenko et al., 1994). NPTII and *neo* mRNA levels were then assessed in the leaves of transplastomic plants. Since NPTII in wild-type DB-containing and mutant DB-containing plants has the exact same protein sequence, protein levels in the plants directly reflect the efficiency of mRNA translation. In case of the *atpB* TCR, mutagenesis of DB reduced protein accumulation to ~4% instead of ~7% (Figure 10 and Table 3). In contrast, mutagenesis of *rbcL* DB had a dramatic effect, reducing NPTII accumulation 35-fold. Thus, DB-ADB interaction is very important for translation of the plastid *rbcL* mRNA, but is less important for translation of the *atpB* mRNA.

We also prepared a third construct set with the *atpB* and *rbcL* leaders, but without the native DB (PrnLatpB-DB, plasmid pHK31, Figure 4B; PrnLrbcL-DB, plasmid pHK35, Figure 4A). The *neo* coding region in these constructs is directly linked to the Prn promoter via a synthetic *NheI* restriction site. The *NheI* restriction site (GCTAGC) is fully complementary to the ADB region (Figure 2B), therefore it was hoped that it would function as a DB sequence. Utility of *NheI* site as an alternative DB could be best judged by NPTII accumulation from the *rbcL* leader, which is highly dependent on DB. High levels of NPTII from the *NheI* construct (4.7%) relative to the mutant DB (0.3%) indicate, that linking the coding region via an *NheI* site provides a suitable DB for expressing foreign polypeptides (Figure 10, Table 3).

TABLE 3  
Levels of NPTII and *neo* mRNA in tobacco leaves

	SD	DB	NPTII (%)	<i>neo</i> mRNA	NPTII/ <i>neo</i> mRNA
Nt-pTNH32-70	+	-	2.10±0.33	41.5	5.06
Nt-pHK30-1D	(+)	wt	7.02±0.82	70.05±12.33	8.85
Nt-pHK31-1C	(+)	s	2.52±0.79	100	2.52
Nt-pHK60-5A	(+)	m	4.03±1.45	91.57±12.76	4.40
Nt-pHK32-2F	-	wt	1.17±0.05	49.33±7.76	2.37
Nt-pHK33-2A	-	s	0.21±0.05	49.55±6.67	0.42
Nt-pHK34-9C	+	wt	10.83±3.84	48.91±22.65	22.14
Nt-pHK35-4A	+	s	4.68±1.84	21.41±7.88	21.86
Nt-pHK64-3A	+	m	0.31±0.15	52.47±4.29	0.59
Nt-pHK36-1C	+	wt	2.17±70.97	68.8	3.15
Nt-pHK37-2D	+	s	2.35±0.05	42.3	5.56

	Nt-pHK38-2E	+	Ec	16.39±3.42	47.59±19.06	34.44
	Nt-pHK39-3B	+	pt	0.16±0.13	13.12±1.27	1.22
	Nt-pHK40-12B	+	s	23.00±5.40	90.27±31.83	25.48
5	Nt-pHK43-1C	(+)	s	0.65±0.28	13.2	4.92

### DISCUSSION

In bacteria, mutagenesis or deletion of the DB  
 10 reduces translation 2- to 34-fold, depending on the  
 individual mRNA (Etchegaray and Inouye, 1999; Faxén et  
 al., 1991; Ito et al., 1993; Mitta et al., 1997;  
 Sprengart et al., 1996). Furthermore, reliance on the DB  
 increases when the SD sequence is removed (Sprengart et  
 15 al., 1996; Wu and Janssen, 1996). In our experiments, no  
 variation was made in the *atpB* or *rbcL* 5'UTR, only  
 sequences downstream of the AUG were altered.  
 Mutagenesis of the *atpB* DB region reduced protein levels  
 -2-fold. Although the *atpB* mRNA does not have a SD  
 20 directly upstream of AUG, we speculate that it probably  
 has an alternate mechanism for translation initiation  
 that reduces its dependence on the DB. Alternatively  
 translation initiation may be facilitated by activator  
 proteins as described for *Chlamydomonas* chloroplasts  
 25 (Rochaix, 1996; Stern et al., 1997). The consequence of  
 DB mutagenesis on *rbcL* translation was a dramatic 35-  
 fold drop in NPTII levels. Accordingly, efficient *rbcL*  
 translation is highly dependent on DB-ADB interactions.  
 Genes in both prokaryotes and eukaryotes show biases in  
 30 the usage of the 61 amino acid codons and have a tRNA  
 population closely matched to the overall codon bias of  
 the resident mRNA population. Incorporation of  
 synonymous minor codons in the coding region may  
 dramatically reduce translation (Makrides, 1996) and  
 35 destabilize the mRNA (Deana et al., 1998). A well-



characterized example for minor codons causing reduced expression in *E. coli* are the AGA/AGG arginine codons recognized by the same tRNA which are present at the frequency of 2.6 and 1.6 per thousand codons.

5 Therefore, we have compared codon usage bias and frequency of triplets per 1000 nucleotides in the wild-type and mutagenized *atpB* and *rbcL* DB regions. Since we studied NPTII accumulation in leaves, the values shown in Figure 12 were calculated for the highly expressed  
10 *rbcL*, *psaA*, *psaB*, *psaC*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE* and *psbF* photosynthetic genes using the Genetics Computer Group (GCG; Madison Wisconsin) codon frequency program. Codon usage bias and triplet frequency is comparable in the wild-type and mutant DB regions of both *atpB* and  
15 *rbcL*. In addition, the mRNAs for the wild-type and mutant DB constructs accumulate at similar levels. Therefore, the dramatic change in NPTII accumulation from the *PrnLrbcL+DBm* promoter in the *Nt-pHK64* line can not be attributed to incorporation of a rare codon in  
20 the mutant DB region.

We have shown here that sequences downstream of the translation initiation codon may dramatically affect mRNA translation. Therefore, silent mutations in the DB region of heterologous proteins may significantly  
25 improve expression in chloroplasts by increasing complementarity of the mRNA with the plastid rRNA penultimate stem structure.

There are significant differences in NPTII accumulation from neo transgenes with different leaders  
30 and the same synthetic DB (Table 3). This indicates that the 5'UTR is an important determinant of translation efficiency. Many data are available supporting the importance of 5'UTR as a target for translational

control in higher plants (Hirose and Sugiura, 1996; Staub and Maliga, 1993; Staub and Maliga, 1994b) and the unicellular alga *Chlamydomonas* (Mayfield et al., 1994; Nickelsen et al., 1999; Sakamoto et al., 1993; Zerges et al., 1997). The data presented herein demonstrate that translation efficiency in plastids is determined by sequences both upstream and downstream of the AUG.

10

**EXAMPLE 2**

**Study of phage T7g10 translation control sequences indicates that the efficient DB in plastids has loose complementarity to ADB**

15 Since the actual ADB sequence is different in plastids and *E. coli*, we anticipated (Sprengart et al., 1996; Etchegaray & Inoyue, 1999) that replacement of the *E. coli* DB with a perfect plastid DB (100% DB-ADB complementarity) would enhance translation in plastids.

20 We choose the phage T7g10 translational control region for the study since it has a well-characterized *E. coli* DB. Three Prn promoter derivatives were constructed. Cassette PrnLT7g10+DB/Ec consists of Prn fused with the native T7g10 TCR containing the *E. coli* DB (plasmid pHK38; Figure 2B, Figure 4A). Cassette PrnLT7g10+DB/pt

25 consists of the Prn promoter, T7g10 leader and the perfect tobacco DB (pHK39; Figure 2B, Figure 4A). Cassette PrnLT7g10-DB has the Prn promoter and T7g10 leader, but lacks the T7g10 DB sequence (pHK40; Figure

30 2B, Figure 4A). The neo coding region in these constructs is directly linked to the Prn promoter via a synthetic *NheI* restriction site. The neo genes in the three expression cassettes were introduced into tobacco

plastids by transformation (Svab and Maliga, 1993; Zoubenko et al., 1994) and the leaves of transplastomic tobacco were tested for NPTII accumulation and mRNA levels (Figures 10, 11; Table 3).

5 Surprisingly, NPTII levels from the heterologous T7g10 TCR were higher (Nt-pHK38; ~16%) than the levels obtained from the *rbcL* TCR (Nt-pHK34; ~11%). We expected that incorporation of the plastid DB with 100% complementarity would further enhance NPTII levels.

10 Instead, we found that plants transformed with the construct having the perfect plastid DB (Nt-pHK39) contained NPTII levels 100-fold lower than the plants expressing NPTII from the *E. coli* TCR (Nt-pHK38; Figures 10; Table 3). This result suggests that, unlike in *E.*

15 *coli*, 100% complementarity reduces, rather than enhances translation efficiency. Indeed, none of the highly expressed plastid genes have a perfect DB sequence (Figure 2A). RNA gel blots shown in Figure 11 indicate that Nt-pHK39 plants with the perfect DB contain ~3-fold

20 less neo mRNA. Therefore, a contributing factor to lower NPTII levels in these plants appears to be a faster mRNA turnover rate. Furthermore, NPTII expressed from the PrnLT7g10 derivatives differ by the DB-encoded amino acids at the N-terminus. Therefore, differential protein

25 turnover rates may be part of the reason for differences in NPTII accumulation. The highest yield of NPTII (23%) was obtained with the synthetic, *NheI*-containing DB cassette.

30

#### DISCUSSION

This example utilizing the *rbcL* translation control regions reveals that sequences downstream of the translation initiation codon may dramatically affect

mRNA translation. Therefore, silent mutations in the DB region of heterologous proteins may significantly improve expression in chloroplasts by increasing complementarity of the mRNA with the plastid rRNA penultimate stem structure. However, it appears that perfect complementarity is undesirable, as it may accelerate mRNA turnover and reduce the rate of translation. This finding highlights differences in the translation machinery of plastids and *E. coli*, in which perfect complementarity enhances translation (Etchegaray and Inouye, 1999; Sprengart et al., 1996). It is possible, however, that shifting the region of complementarity relative to AUG or targeting a slightly different region of the penultimate stem may facilitate highly efficient translation of mRNAs with a perfectly matched DB.

The T7g10 constructs have one or two relatively rare AGC serine codons (4.7 per 1000, Figure 12), one of which is encoded in the NheI site. This codon is present in the Nt-pHK38 and Nt-pHK40 plants, which contain the highest levels of NPTII. Further improvement may be expected by replacing the AGC with an AGT serine codon.

### EXAMPLE 3

**The *clpP*, *psbB* and *psbA* TCRs have distinct expression characteristics**

NPTII accumulation was studied in transplastomic tobacco carrying the *PrrnLclpP* promoter derivatives. The *PrrnLclpP*+DBwt (Nt-pHK32-2F) and *PrrnLclpP*-DB (Nt-pHK33-2A) plants accumulate 1.2% and 0.2% NPTII in their leaves (Figure 10; Table 3). We have found that over-

expression of *clpP* 5'-UTR causes a mutant phenotype manifested as pale green leaf color and slower growth. This phenotype is normalized in older plants. We assume that the primary cause of this mutant phenotype is the lack of ClpP protein, the *clpP* gene product. This mutant phenotype is absent in plants transformed with other 5'UTRs. Therefore we believe, that the mutant phenotype is attributable to competition for a *clpP*-specific nuclear factor. The *clpP* gene has two introns. Preliminary RNA gel blot analysis reveals reduced levels of mature, monocistronic *clpP* mRNA (~30% of wild-type) and accumulation of intron I-containing *clpP* pre-mRNA in the pale-green leaves. Normalization of phenotype coincides with increase of translatable monocistronic *clpP* mRNA to wild type levels. Over-expression of *clpP* 5'UTR therefore may interfere with splicing of *clpP* pre-mRNA.

NPTII accumulation was also studied in transplastomic tobacco carrying the *PrrnLpsbB* promoter derivatives. The *PrrnL psbB+DBwt* (Nt-pHK36-1C) and *PrrnL psbB -DB* (Nt-pHK37-2D) plants accumulate 2.2% and 2.4% NPTII in their leaves (Figure 10; Table 3). Thus, the synthetic DB sequence in case of the *psbB* TCR efficiently replaces the native DB sequence. Conversely, it may rely on an alternative mechanism for translation initiation.

The *Prrn* promoter constructs with the *psbA* leader were obtained as described. However, we have been able to introduce only one of them, *PrrnLpsbA-DB(+GC)* into tobacco plastids in line Nt-pHK43-1C. The Nt-pHK43-1C plants accumulate NPTII at a relatively low level (0.6%; Figure 10, Table 3). It is conceivable that the lack of success in introducing the +DB construct is due to the

dramatically elevated expression level of NPTII which is toxic to the plants.

5

#### DISCUSSION

NPTII levels obtained from PrnLclpP+DBwt (Nt-pHK32-2F) promoter are relatively low, only 1.2% of the total soluble protein. However, this promoter is  
10 desirable for driving expression of selectable marker genes, as the recovery of transplastomic clones is relatively efficient when the *neo* gene is expressed from this promoter, as shown in Example 4. Expression of *neo* from the PrnLclpP+DBwt promoter does not cause a mutant  
15 phenotype in tissue culture. Thus, it is suitable to drive the expression of marker genes, so long as the marker gene is subsequently removed. It appears that competition for a nuclear-encoded factor required for processing the *clpP* introns gives rise to the reduced  
20 expression observed. This intron is absent in the *clpP* genes in the monocots rice (Hiratsuka et al., 1989) and maize (Maier et al., 1995). The PrnLclpP+DBwt promoter therefore may be used to advantage in the transformation of monocots. Furthermore, the level of the trans-factor  
25 required for *clpP* intron processing is likely to be expressed at different levels in dicots. We anticipate therefore, that expression of the *clpP* TCR will have no undesirable consequences in other dicot species. It is also possible that the phenotypic consequences of  
30 expressing the *clpP* TCR in plastids is a property of the tobacco line, *N. tabacum* cv. Petit Havana utilized herein and is absent in other tobacco lines. This would

make the *clpP* gene TCR a desirable expression tool in both monocots and dicots.

Both *psbB* leader derivatives accumulate NPTII at comparable levels (2.2% and 2.4%, respectively; Table 3). This 5' regulatory region is a good alternative to the most commonly used *rbcL* leader when protein accumulation is required in the ~2% range.

In the past, the *psbA* promoter and leader construct yielded relatively high levels of expression in leaves (2.5% GUS; Staub and Maliga, 1993). Yet these constructs did not contain *psbA* DB elements. The present invention describes the generation of chimeric promoters that are suitable to obtain high-level protein expression while elucidating the regulatory role played by DB sequences. *Prn* is the strongest known promoter in plastids and consequently provides for high levels of NPTII translation. These elevated levels of NPTII can be toxic to the plant and therefore it is difficult to obtain transplastomic lines with the highest prospective levels of NPTII. An alternative approach involves operably linking the *psbA* leader to a relatively weak promoter. This approach may generate cassettes which are suitable for obtaining relatively high levels of protein accumulation from relatively low levels of mRNA.

25

#### EXAMPLE 4

##### NPTII accumulation in roots and seeds

Posttranscriptional regulation is an important mechanism of plastid gene expression (Rochaix, 1996; Stem et al., 1997). Therefore, we expected that NPTII accumulation may be tissue-specific due to regulation of

30

gene expression at the level of mRNA translation. Thus, NPTII accumulation was tested in roots and seeds.

Testing of NPTII accumulation in roots was carried out with a subset of transplastomic lines (Table 4).

5 Roots for protein extraction were collected from plants grown in liquid MS salt medium (3% sucrose) in sterile cultures incubated on a shaker to facilitate aeration. Protein was extracted from the roots with the leaf protocol and tested for NPTII accumulation (Figure 13  
10 A). The highest level of NPTII, 0.75%, is found in the roots of plants expressing NPTII from the *clpP* TCR (PrnnLclpP+DBwt construct; pHK32). The second highest value, 0.3%, was found in the roots of plants transformed with plasmid pHK38 expressing NPTII from the  
15 T7g10 TCR (PrnnLT7g10+DB/Ec promoter). The level of NPTII was about the same, approximately 0.1 %, in roots expressing the recombinant protein from the *atpB* and *rbcL* TCR in pHK30- and pHK34- transformed plants.

Since plastids in the roots are smaller than in  
20 leaves, we expected lower levels of NPTII accumulation in the roots than in the leaves. This was true for all the tested roots, except those of the Nt-pHK32 plants. Interestingly, NPTII from the *clpP* TCR accumulated at almost the same level in the roots (0.75%, Table 4) as  
25 in the leaves (approximately 1%, Table 3). This is likely attributable to high levels of the *neo* mRNA in the roots (Figure 13B). Since the *clpP* leader includes the minimal P*clpP*-53 promoter (Sriraman et al., 1998a; NAR 26: 4874) we speculate, that the relatively high  
30 mRNA levels are due to activation of P*clpP*-53 in roots. High levels of expression make the *clpP* leader a desirable TCR for protein expression in roots.



The T7g10 leader (pHK38) was the most efficient in roots from which the most NPTII accumulated relative to the mRNA (Table 4). Although in the Nt-pHK38 plants, the neo mRNA was 7-times less than in the Nt-pHK32 plants, NPTII levels were almost as high (approximately 0.30% compared to 0.75%) as in the plastids with the clpP TCR (pHK32). High level NPTII accumulation from the T7g10 TCR in leaves (pHK38, pHK40; Table 3) and in roots (pHK38; Table 4) indicates the general utility of the phage T7g10 translation control region for protein expression in plastids.

Protein accumulation was also studied in seeds harvested from the transgenic plants (Figure 14). Protein levels were 0.05% in plants transformed with pHK32 (clpP TCR), and approximately 0.01% in plants transformed with plasmid pHK30 (atpB TCR). No NPTII was detectable in plants in which neo was introduced in the rbcL TCR-construct (plasmid pHK34), indicating differential protein accumulation which is dependent on the choice of the TCR.

Table 4.  
Levels of NPTII and neo mRNA in tobacco roots

Strain	NPTII (%)	neo mRNA (%)	NPTII/neo mRNA $\times 10^3$
Nt-pHK30-1D	0.14 $\pm$ 0.05	33.7	4.2
Nt-pHK32-2F	0.75 $\pm$ 0.35	100	7.5
Nt-pHK34-9C	0.12 $\pm$ 0.03	23.5	5.1
Nt-pHK38-2E	0.31 $\pm$ 0.04	13.4	23.1

5

**EXAMPLE 5**

**High-level NPTII expression facilitates efficient  
recovery of transplastomic lines by selection for  
kanamycin resistance**

10

The plastid genome of higher plants is a 120-kb to 160-kb double-stranded DNA which is present in 1,900 to 50,000 copies per leaf cell (Bendich, 1987). To obtain genetically stable transplastomic lines every one of the  
15 plastid genome copies (ptDNA) should be uniformly altered in a plant. Since integration of foreign DNA always occurs by homologous recombination, plastid transformation vectors contain segments of the plastid genome to target insertions at specific locations.  
20 Useful, non-selectable genes are cloned next to the selectable marker genes, which are then introduced into the plastid genome by linkage to the selectable marker gene (Maliga, 1993). Transforming DNA is introduced into plastids by the biolistic process (Svab et al., 1990;  
25 Svab and Maliga, 1993) or PEG treatment (Golds et al., 1993; O'Neil et al., 1993). Elimination of wild-type genome copies occurs during repeated cell divisions on a selective medium. The success of transformation depends on the success of selective amplification of the few  
30 initially transformed genome copies. Therefore the choice of the antibiotic used for the selective amplification of transformed genome copies and the mechanism by which the plant cells are protected from

antibiotic action is a critical parameter to be considered for successful generation of homoplasmic plants.

5 The most commonly used antibiotic for the selection of transplastomic lines is spectinomycin, an inhibitor of protein synthesis on plastid ribosomes. Initially, plastid transformation in tobacco was carried out by selection for resistance based on mutations in the plastid 16S rRNA (Svab et al., 1990). Selection was  
10 inefficient, yielding about one transplastomic clone per 50 bombarded samples, probably because the 16S rRNA based mutation is recessive. Recovery of transplastomic lines was enhanced ~100-fold by selection for a dominant marker, spectinomycin resistance based on inactivation  
15 by aminoglycoside 3" adenylyltransferase encoded in a chimeric *aadA* gene (Svab and Maliga, 1993). In addition to tobacco, selection for spectinomycin resistance (*aadA*) could be applied to recover transplastomic lines in *Arabidopsis* and potato. The *aadA* gene in plants  
20 confers resistance to both spectinomycin and streptomycin. Selection for streptomycin resistance was used for plastid transformation in rice, a species resistant to spectinomycin, after bombardment with a chimeric *aadA* gene. See Example 8.

25 The need for an alternative marker gene for plastid manipulation has led to testing kanamycin resistance as a selective marker. A chimeric *neo* (*kan*) gene, encoding neomycin phosphotransferase, was suitable to recover transplastomic tobacco lines. However, recovery of  
30 transplastomic lines was relatively inefficient, yielding only one transplastomic line in ~25 bombarded leaf samples. Furthermore, for every plastid transformation event ~25 to 50 kanamycin resistant lines

were obtained in which integration of the plastid *neo* construct into the nuclear genome resulted in kanamycin resistance (Carrer et al., 1993). We report here that the efficiency of recovering transplastomic clones is significantly improved when transforming tobacco chloroplasts with a new *neo* gene expressed from a promoter with the *atpB* and *clpP* translation control region. The number of nuclear transformation events is reduced using the cassettes of the present invention. These improvements make the new *neo* gene a practical tool for plastid genome manipulations.

#### DISCUSSION

The chimeric *neo* genes described in Examples 1-4 were introduced into plastids by selection for the linked spectinomycin resistance (*aadA*) gene as their suitability for directly selecting transplastomic lines was unknown. The transplastomic lines listed in Table 3 were then tested for resistance to kanamycin by their ability to proliferate on a medium containing 50 mg/L kanamycin. The RMOP medium used for testing induces formation of green callus and shoot regeneration in the absence of kanamycin. The tissue culture procedures utilized for this example are described in references Carrer et al., 1993 and Carrer and Maliga, 1995.

On the selective kanamycin medium only scanty, white callus forms from wild-type leaf section. Formation of green callus and shoots from leaf section of plants transformed with pHK plasmids in Table 3 indicates that accumulation of NPTII confers kanamycin resistance. We set out to test if transplastomic clones can be directly selected by kanamycin resistance after bombardment with

plasmids pHK30 and pHK32. The results are summarized in Table 5.

Bombardment of 25 tobacco leaves with plasmid pHK30 yielded 45 kanamycin resistant lines on a medium  
5 containing 50 mg/L kanamycin. Transplastomic *neo* lines are expected to be resistant to much higher levels, 500 mg/L of kanamycin (Carrer et al., 1993). In addition, in plasmid pHK30 the *neo* gene is physically linked to a spectinomycin resistance (*aadA*) gene. Spectinomycin  
10 resistance is manifested as kanamycin resistance: sensitive leaf sections form white callus and no shoots whereas resistant leaf sections form green callus and shoots on a selective medium (500 mg/L) RMOP medium. We assumed therefore, that all transplastomic lines  
15 should be resistant to both 500 mg/L of kanamycin and 500 mg/L spectinomycin (Carrer and Maliga, 1995). When applying this test we found that 22 of the 45 lines meet these criteria. Digestion of the plastid DNA with the EcoRI restriction enzyme and probing with the plastid  
20 targeting region should detect 3.1-kb fragment in the wild-type and a 4.2-kb and 1.2-kb fragment in transplastomic lines (Figure 15A). DNA gel blot analysis of seven of the kanamycin-spectinomycin resistant lines confirmed integration of both transgenes into the  
25 plastid genome (Figure 15B). Therefore, we assume that all 22 kanamycin-spectinomycin lines are transplastomic (Table 5).

Bombardment of 30 tobacco leaves with plasmid pHK32 yielded 28 kanamycin resistant lines on a medium  
30 containing 50 mg/L kanamycin. We have identified 11 double-resistant lines by testing these on a medium containing 500 mg/L of kanamycin and 500 mg/L spectinomycin. All six tested were transplastomic by DNA

gel blot analysis (Figure 15B), therefore we believe that all eleven are transplastomic (Table 5).

5

TABLE 5						
SELECTION OF TRANSPLASTOMIC TOBACCO						
CLONES BY KANAMYCIN RESISTANCE						
Vector	No.	Kan. Res.	Kan. Res.	Kan. Res.	Transplastomic	
	leaves	50 mg/L	500 mg/L	500 mg/L		
				Spec. Res.		
				500 mg/L		
pTNH32	29	59	7			0
	50 <sup>a</sup>	52				2
	25 <sup>a</sup>	47	4			1
pHK30	25	45		22		22
pHK32	30	28		11		11

15

20

---

(<sup>a</sup>Carrer et al., 1993)

25

#### DISCUSSION

30

35

Plastid transformation efficiency should be comparable, if we target the same region of the plastid genome for insertion, use similar size targeting sequences and the same method of DNA delivery. Therefore, lower transformation efficiencies obtained by selection for kanamycin resistance with the old chimeric neo genes was likely due to the lack of recovery of transplastomic clones by selection. We have found that transformation with neo genes expressed from the

PrnLatpB+DBwt and PrnLclpP+DBwt promoters is as efficient as with the aadA gene. This is a significant technical advance, and will facilitate plastid transformation in crops, in which the regenerable tissues contain non-green plastids. Most important targets are the non-green plastids of cereal crops. Kanamycin selection is widely used to obtain transgenic lines after transformation with chimeric neo genes in dicots. However, kanamycin is an undesirable selective agent in monocots such as cereal tissue cultures. However, NPTII also inactivates paromomycin, which may be used to recover nuclear gene transformants at an extremely high efficiency in cereals. See for example, PCT application WO99/05296.

#### EXAMPLE 6

**Bacterial bar gene expression in tobacco plastids confers resistance to the herbicide phosphinothricin**

Bialaphos, a non-selective herbicide, is a tripeptide composed of two L-alanine residues and an analog of glutamic acid known as phosphinothricin (PPT). While PPT is an inhibitor of glutamine synthetase in both plants and bacteria, the intact tripeptide has little or no inhibitory effect in vitro. Bialaphos is toxic for bacteria and plants, as intracellular peptidases remove the alanine residues and release active PPT. Bialaphos is produced by *Streptomyces hygroscopicus*. The bacterium is protected from phosphinothricin toxicity by phosphinothricin acetyltransferase (PAT), the bar gene product. This enzyme acetylates phosphinothricin or demethylphosphinothricin (Thompson et al., 1987). PPT resistant crops have been obtained by expressing the S.

*hygroscopicus* bar gene in the plant nucleus. Herbicide resistant lines were obtained by direct selection for PPT resistance in culture after *Agrobacterium tumefaciens*-mediated DNA delivery in tobacco, potato, *Brassica napus* and *Brassica oleracea* (De Block et al., 1987, 1989). Biolistic DNA delivery of chimeric bar genes has been employed to obtain PPT resistant maize (Spencer et al., 1990), rice (Cao, et al, 1992) and *Arabidopsis thaliana* (Sawaskaki et al., 1994).

Construction of transplastomic tobacco plants, in which PPT resistance is based on the expression of bar from *S. hygroscopicus* in plastids is described in the present example. The vectors utilized to express the bar gene contain an exemplary chimeric 5' regulatory region as set forth in the previous examples. The following material and methods facilitate the practice of this aspect of the present invention.

#### Construction of plastid bar gene

A NcoI/XbaI bar gene fragment was generated by PCR amplification using plasmid of pDM302 (Cao et al., 1992) with the following primers:

P1, 5'-AAACCATGGCACCACAAACAGAGAGCCCAGAACGACGCCC-3';

P2, 5'-AAAATCTAGATCATCAGATCTCGGTGACG-3'.

The ends of the PCR fragment were blunt ended by treatment with the Klenow Fragment of DNA polymerase I. The fragment was then ligated into the EcoRV site of pBluescript II KS+ (Stratagene, La Jolla, CA) to create plasmid pJEK3. Sequence analysis of pJEK3 plasmid DNA revealed that the XbaI site we intended to create through PCR amplification of pDM302 is absent. See Figure 19. The bar gene has the two translation



termination codons followed by vector sequences. The last 20 bp of pJEK3 are:

CCCGTCACCGAGATCTGATGATcgaattcctgcagcccggggatccactagttct

aga. The bar sequences are in capital (stop codons

underlined), the vector sequences are in lower case

(XbaI site underlined). Since there is an XbaI site

present in the vector 40 bp from the intended XbaI site,

it was not necessary to repair this error. The NcoI-XbaI

fragment from plasmid pJEK3 was ligated into NcoI-XbaI

digested pGS104 plasmid (Serino and Maliga, 1997) to

generate plasmid pJEK6. Plasmid pGS104 carries a Prn-

TrbcL expression cassette in a pPRV111B plastid

transformation vector. A map of the plastid targeting

region of plasmid pJEK6 is shown in Figure 16A.

#### Plastid transformation and plant regeneration

Tobacco (*Nicotiana tabacum* cv. Petit Havana)

plants were grown aseptically on agar-solidified medium

containing MS salts (Murashige and Skoog, 1962) and

sucrose (30g/l). Leaves were placed abaxial side up on

RMOP media for bombardment. The RMOP medium consists of

MS salts, N6-benzyladenine (1mg/l), 1-naphthaleneacetic

acid (0.1 mg/l), thymine (1mg/l), inositol (100 mg/l),

agar (6g/l), pH 5.8, and sucrose (30g/l). The DNA was

introduced into chloroplasts on the surface of 1 $\mu$ m

tungsten particles using the DuPont PDS1000He Biolistic

gun (Maliga 1995). Spectinomycin resistant clones were

selected on RMOP medium containing 500  $\mu$ g/ml

spectinomycin dihydrochloride. Resistant shoots were

regenerated on the same selective medium and rooted on

MS agar medium (Svab and Maliga, 1993). The

independently transformed lines are designated by the

transforming plasmid (pJEK6) and a serial number, for example pJEK6-2, pJEK6-5. Plants regenerated from the same transformed line are distinguished by letters, for example pJEK6-2A, pJEK6-2B.

5

#### Southern Blot Analysis

Total cellular DNA was isolated from wild-type and transgenic spectinomycin resistant plants with CTAB (Saghai-Maroo et al., 1984). The DNA was digested with the Sma I and BglII restriction endonucleases, separated on a 0.7% agarose gel and blotted onto a Hybond-N nylon membrane (Amersham, Arlington Heights, IL) by a pressure blotter. The membrane was hybridized overnight with an ApaI/ BamHI fragment labeled with ( $\alpha$ -  $^{32}$ P )dCTP using a dCTP DNA Labeling Beads Kit (Pharmacia Inc, Piscataway, NJ). The membrane was washed 2 times with 0.1X SSPE, 0.2X SDS at 55°C for 30 minutes. Film was exposed to the membrane for 30 minutes at room temperature.

#### PAT Assay

The PAT assay was performed as described by Spencer et. al. (1990). Leaf tissue (100 mg) from wild type tobacco (wt), transgenic Nt-pDM307-10 tobacco (a line transformed with the nuclear bar gene in plasmid pDM307; Cao et al., 1992), and plastid bar gene transformants was homogenized in 1 volume of extraction buffer (10 mM  $\text{Na}_2\text{HPO}_4$ , 10 mM NaCl). The supernatant was collected after spinning in a microfuge for 10 minutes. Protein (25 mg) was added to 1 mg/ml PPT and  $^{14}\text{C}$ -labeled Acetyl CoA. The reaction was incubated at 37°C for 30 minutes and the entire reaction was spotted onto a TLC plate. Ascending

chromatography was performed in a 3:2 mixture of 1-propanol and  $\text{NH}_4\text{OH}$ . Film was exposed to the TLC plate overnight at room temperature.

5

#### Herbicide Application

Wild type and transgenic plants were sprayed with 5 ml of a 2% solution of Liberty (AgrEvo, Wilmington, DE) with an aerosol sprayer.

#### RESULTS AND DISCUSSION

First the bacterial bar gene was converted into a plastid gene by cloning the bar coding region into a plastid expression cassette. This cassette consists of an engineered plastid rRNA operon promoter (Prn) and TrbcL and the 3' UTR of the plastid rbcL gene for stabilization of the mRNA. The plastid bar gene was then cloned into the plastid transformation vector to yield plasmid pJEK6, and introduced into plastids on the surface of microscopic tungsten particles. The bar gene integrated into the plastid genome by two homologous recombination events via the plastid targeting sequences, as shown in Figure 16A. Selection for the linked aadA (spectinomycin resistance) gene on spectinomycin-containing medium eventually yielded cells which carried a uniformly transformed plastid genome population, which were then regenerated into plants.

Integration of bar and aadA was verified by DNA gel blot analysis. Total cellular DNA of wild-type and transplastomic plants was digested with the SmaI and

BglIII restriction enzymes and probed with the 2.9-kb  
ApaI-BamHI plastid targeting fragment of *N. tabacum*  
(Figure 16B). The two fragments that were expected for  
the transgenic plants, 3.3 kb and 1.9 kb, were present  
5 in each of the transplastomic samples shown in Figure  
16B. Absence of the 2.9 kb wild type fragment indicated,  
that by the time these plants have been regenerated, the  
wild-type plastid genome copies have been diluted out on  
the selective medium.

10 To determine if the plastid bar gene has been  
expressed, leaf extracts were assayed for  
phosphinothricin acetyltransferase (PAT) activity.  
Conversion of PPT into acetyl-PPT indicated PAT activity  
in each of the tested transplastomic lines. Data in  
15 Figure 17 are shown for the transplastomic lines Nt-  
pJEK6-2D, Nt-pJEK6-5A and Nt-pJEK6-13B. Interestingly,  
PAT activity was significantly (>>10-fold) higher when  
bar was expressed in the plastids, as compared to the  
bar gene expressed from the cauliflower mosaic virus 35S  
20 promoter in the nucleus of the Nt-pDM307-10 plant.

PAT expression confers resistance to PPT in tissue  
culture and in the greenhouse. When wild type leaf  
sections are grown in tissue culture, 10 mg/L PPT  
completely blocks callus proliferation. This same PPT  
25 concentration is suitable for the selection of nuclear  
transformants after bombardment with the nuclear bar  
construct in plasmid pDM307. Leaf sections of plants  
expressing bar in plastids show resistance in the  
presence of up to 100 mg/L PPT in the culture medium. We  
30 have tested PPT resistance in the greenhouse, spraying  
wild-type and transplastomic plants with Liberty, a  
commercial formulation of PPT, at the recommended field  
dose of 2%. As shown in Figure 18A, 13 days after the

treatment, the wild type plants were dead while the transgenic plants thrived. Since then the sprayed plants have flowered and set seed. Figure 18B shows maternal inheritance of PPT resistance. Lack of plastid pollen transmission results in a lack of herbicide resistance in progeny pollinated with transgenic pollen. The bacterial bar gene has a high G + C content (68.3%; Genbank Accession No. X17220), while plastid genes have a relatively high A + T content; for example the G + C content of the highly expressed psbA and rbcL genes is 42.7% and 43.7%, respectively (Genbank Accession No. 200044). Differences in the G + C content are also reflected in the codon usage biases. Interestingly, data presented here indicate that expression of bar from *S. hygroscopicus* is sufficiently high to confer resistance to field levels of the non-selective herbicide PPT. Furthermore, the PAT enzyme levels obtained in the transplastomic lines are significantly higher than those observed in the nuclear transformant. Therefore, further improvement of the expression levels may be obtained by optimizing the codon usage for plastids as set forth in Example 7.

Advantages of incorporating bar in the plastid genome are containment of herbicide resistance due to the lack of pollen transmission in most crops. Furthermore, the lack of genetic segregation would simplify back-crossing for the introduction of herbicide resistance into additional breeding lines.

30

#### EXAMPLE 7

##### A Synthetic bar gene Improves Containment and Enhances Expression in Plastids

The bacterial *bar* gene was introduced into the tobacco plastid genome by transformation with plasmid pJEK6, as described above in Example 6. In plasmid pJEK6  
5 *bar* is expressed in a cassette consisting of the Prn(L)rbcL(S) promoter and TrbcL transcription terminator. This plasmid conferred PPT resistance to plants grown in the presence of PPT in the tissue culture medium, but direct selection for transformed  
10 lines was not possible. Although the PAT levels in homoplastomic leaves was high, the amount of PAT produced by the few pJEK6 *bar* copies during the early stage of plastid transformation was probably insufficient to protect the entire cell.

15 To improve *bar* expression in plastids a synthetic gene was created. The codon usage was modified to mimic that of the average tobacco photosynthetic plastid gene. Changing the codon usage lead to a lowered GC content characteristic of higher plant plastid genes. To assist  
20 with cloning, restriction enzyme recognition sequences were removed and added as necessary. Codon usage frequency in bacteria reflects relative tRNA abundance: frequent use of codons for rare tRNAs may significantly reduce translation efficiency. We hoped that  
25 differential codon usage in plastids and bacteria would reduce or prevent expression of the synthetic gene in bacteria, thereby reducing the danger of horizontal gene transfer to microorganisms. We also hoped that improved *bar* expression in our novel promoter cassettes will  
30 allow direct selection of plastid transformants on PPT-containing medium.

#### Materials and Methods for Example 7

Codon comparisons of photosynthetic (*rbcL*, *psaA*, *psaB*, *psaC*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*) plastid genes were compiled using GCG (Genetics Computer Group, Madison, WI). DNA mutations were then introduced into the bacterial *bar* gene making its codon usage more similar to plastid genes, while removing several restriction enzyme sites that could interfere with cloning. See Figure 28. The synthetic *bar* gene (*s-bar*) was obtained by single-step assembly of the entire *s-bar* gene from 28 oligonucleotides (one 44 nt primer, one 30 nt primer and twenty-six 40 nt primers) using PCR (Stemmer et al., 1995). The top and bottom strands of the primers overlap with each other by 20 nucleotides. *NcoI* and *NheI* sites were added at the 5' end and a *XbaI* site was added at the 3' end through PCR amplification. To obtain the complete *s-bar* gene, a small aliquot of the assembly PCR product was amplified using primers 1A and 14B. Unchanged nucleotides are in upper case, altered nucleotides are in lower case in the primers listed below.

Primer 1A ccATGgctAGCCCAGAAaGAaGaCCGGCCGAtATtaGaCG  
 Primer 1B GCATaTCaGcTcTcGTaGCACGtCtaATaTCGGCCGGtCt  
 Primer 2A TGcTACaGAaGcTgAtATGCCaGCaGtTtGtACaATCGTt  
 Primer 2B CTTGtTtTcTATaTAaTGGTTaACGATtGTaCAaACTGcTG  
 Primer 3A AACCAtTAtATaGAaACAAGtACaGTaAACTTtaGaActG  
 Primer 3B tTcTtGaGGTTcTtTGaGGtTCaGTtCtaAAGTTtActGTa  
 Primer 4A AaCcTCAaGAACcTCAaGAaTGGACtGAtGAtCTaGTCCG  
 Primer 4B AaGGATAGCGCTcTcGTAGACGGACtAGaTCaTCaGTCCA  
 Primer 5A TCTaCGaGAGCGCTATCctTGGCTtGTaGCaGAaGtTGAC  
 Primer 5B GCGATaCCaGcTACTTCaCCGTCaACTTcTGcTACaAGCC  
 Primer 6A GGtGAaGTaGcTGGtATCGCaTAtGCGGGCCcTtTGAAGG  
 Primer 6B CCAaTCaTATGCaTtTcTtGCCTTCCAaGGGCCCGCaTAT  
 Primer 7A CAaGaAAtGCaTATGAtTGGACaGcTGAaTCaACTGtTtTA

Primer 7B GtTGaTGaCGtGGtGAaACGTAAaCaGtTgATTCaGctGT  
 Primer 8A CGTtTCaCCaCGtCatCAaCGtACaGGACTtGGtTctAct  
 Primer 8B TTCAGtAGaTGtGTaTAtAGaGTaGAaCCaAGtCCtGTaC  
 Primer 9A CTaTAtACaCatCTaCTGAAaTcttTGGAGGCACAaGGtT  
 5 Primer 9B aACAGctACaACaCTCTTaAAaCCtTGTGCCTCCAaaGAt  
 Primer10A TtAAGAGtGTtGTaGCTGTtATaGGatTGCCTAAatGAtCC  
 Primer10B CtTCaTGCATGCGtACaCtTGGaTCaTTaGGCAatCCtAT  
 Primer11A aAGtGTaCGCATGCATGAaGctCTaGGATATGctCCaaGa  
 Primer11B CCtGCaGCCCtCAaCATaCCtCttGGaGCATATCCtAGaG  
 10 Primer12A GGtATGtTGaGGGctGCaGGtTTCAAaCatGGaAACTGGC  
 Primer12B tTGCCaAAAACctACaTCATGCCAGTTtCCaTGTtTGAAa  
 Primer13A ATGAtGTaGGTTTtTGGCAaCTtGAtTTCAGtCTaCCaGT  
 Primer13B GtAGaActGGACGaGGaGGTACTGGtAGaCTGAAaTCaAG  
 Primer14A ACctCctCGTCCaGtTcTaCCaGtTActGAGATCTGATGA  
 15 Primer14B tctagaTCATCAGATCTCaGtAACTG

The amplified *s-bar* coding region was then cloned into a pBSIIKS+ plasmid (Stratagene, La Jolla, CA) and sequenced (Figure 20A). The *s-bar* gene was cloned into cassettes with the chimeric PrnLatpB+DBwt,

20 PrnLrbcL+DBwt and PrnLT7g10+DB/Ec promoters. Table 6 sets forth the plasmids used in the practice of this example.

Table6. Plasmids with *bar* genes.

25

30

Plasmid	Promoter	<i>bar</i>	3'UTR	Vector
pK05		synthetic ( <i>s-bar</i> )		pBSIIKS+
pK03	PrnLatpB+DBwt	synthetic ( <i>s-bar</i> )	TrbcL	pPRV111B
pK08	PrnLrbcL+DBwt	synthetic ( <i>s-bar</i> )	TrbcL	pPRV111A
pK017	PrnLT7g10+DB/ Ec	synthetic ( <i>s-bar</i> )	TrbcL	pPRV111B
pK012	PrnLrbcL+DBwt	bacterial ( <i>bar</i> )	TrbcL	pPRV111A



To provide a suitable cloning site at 3'-end of the bacterial *bar* gene, the *EagI*/*BglII* fragment of *s-bar* was replaced with the cognate fragment of the bacterial *bar* coding region. Such a bacterial *bar* gene is incorporated in plasmid pK012 (Figure 21). In plasmid pK012 the first 22 nucleotides of the bacterial *bar* coding region are replaced with nucleotides from the *s-bar*.

10

## RESULTS

The engineered bacterial *bar* gene in pJEK6 is expressed both in *E. coli* and plants, as shown in the previous example. We were interested to test if modification of the codon affects expression of the *s-bar* gene in plastids and in *E. coli*. In *E. coli*, *s-bar* expression was determined by measuring PAT activity. Extracts were prepared from bacteria carrying plasmids pK03 and pK08 expressing *s-bar* from the *PrrnLatpB*+DBwt and *PrrnLrbcl*+DBwt promoters, respectively. The radioactive assay did not detect any activity, although extracts from bacteria transformed with plasmids pJEK6 and pK012 carrying the bacterial *bar* genes gave strong signals (Figure 22A). In plasmid pK012 the first 22 nucleotides of the bacterial *bar* coding region are replaced with nucleotides from the *s-bar*. Therefore, lack of expression from the *s-bar* in *E. coli* is not due to changes within the first 22 nucleotides.

The *s-bar* was also introduced into plastids by transformation with vector pK03. Extracts were prepared from pK03- and pJEK6-transformed tobacco plants, which carry the *s-bar* and *bar* genes, respectively. Extracts from both types of plants contained significant PAT

30

activity (Figure 22B). Therefore, the synthetic *bar* is expressed in plastids but not in *E. coli*.

Changing the *bar* gene codon usage abrogated expression of the gene in *E. coli*. This is likely due to the introduction of the rare AGA and AGG arginine codons in the *s-bar* coding region. The triplet frequency per thousand nucleotides for AGA and AGG is the lowest in *E. coli*, reflecting low abundance of the tRNA required for translation of these codons. The minor arginine tRNA<sup>Arg (AGG/AGA)</sup> has been shown to be a limiting factor in the bacterial expression of several mammalian genes. The coexpression of *ArgU* (*dnaY*) gene that encodes for tRNA<sup>Arg (AGG/AGA)</sup> resulted in high level production of the target protein (Makrides 1996). The bacterial *bar* gene has 14 arginine codons, none of which are the rare AGA/AGG codons. The *s-bar* gene has five of them, three of which are located within the first 25 codons. Therefore, the likely explanation for the lack of *s-bar* expression in *E. coli* is introduction of the rare AGA and AGG arginine codons in the *s-bar* coding region.

There are proteins, which are toxic to *E. coli* but their expression is desirable in plastid to which it is not toxic. Engineering of these proteins in *E. coli* poses a problem, since the commonly used PEP plastid promoters are active in *E. coli*, thus the gene will be transcribed and the mRNA translated. Incorporation of minor codons in the coding region will prevent translation of these proteins in *E. coli*. Particularly useful in this regard is conversion of arginine codons to AGA/AGG. If no arginine is present in the N-terminal region, an N-terminal fusion may be designed containing multiple AGA/AGG codons to prevent translation of the mRNA.

Plants under field conditions are associated with microbes living in the soil, on the leaves and inside the plants. Gene flow from plastids to these microorganisms has not been shown. However, it would be an added safety measure to incorporate codons in plastid genes, which are rare in the target microorganisms, but are efficiently translated in plastids. Incorporation of AGA/AGG codons into the selective marker genes and the genes of interest will prevent transfer of genes from plants to microbes, which lack the capacity to efficiently translate the AGA/AGG codons. In case of specific plant-microbe associations, based on differences in codon usage preferences genes could be designed which would be expressed in plastids but not in microbes.

Attempts to directly select transplastomic clones after bombardment with the *s-bar* constructs so far has failed. The *s-bar* coding region in Figure 20A contains frequent and rare codons in proportions characteristic of plastid genes. It is possible, that relatively rare codons in a specific context at a critical stage will prevent recovery of plastid transformation events. Examples for tissue-specific translation of mRNAs dependent on tRNA availability are known (Zhou et al., 1999). Therefore, we designed a second synthetic bar gene, *S2-bar*, containing only frequent codons (Figure 20B). Plastid transformation with the *s2-bar* will enable direct selection of plastid transformation events by PPT resistance.

30

#### EXAMPLE 8

FLUORESCENT ANTIBIOTIC RESISTANCE MARKER FOR FACILE IDENTIFICATION OF TRANSPLASTOMIC CLONES IN TOBACCO AND RICE

Plastid transformation in higher plants is accomplished through a gradual process, during which all the 300-10,000 plastid genome copies are uniformly altered. Antibiotic resistance genes incorporated in the plastid genome facilitate maintenance of transplastomes during this process. Given the high number of plastid genome copies in a cell, transformation unavoidably yields chimeric tissues, in which the transplastomic cells need to be identified and regenerated into plants. In chimeric tissue, antibiotic resistance is not cell autonomous: transplastomic and wild-type sectors both are green due to phenotypic masking by the transgenic cells. Novel genes encoding FLARE-S, a fluorescent antibiotic resistance enzyme conferring resistance to spectinomycin and streptomycin, which were obtained by translationally fusing aminoglycoside 3'-adenylyltransferase [AAD] with the *Aequorea victoria* green fluorescent protein (GFP) are provided in the present example. FLARE-S facilitates distinction of transplastomic and wild-type sectors in the chimeric tissue, thereby significantly reducing the time and effort required to obtain genetically stable transplastomic lines. The utility of FLARE-S to select for plastid transformation events was shown by tracking segregation of transplastomic and wild-type plastids in tobacco and rice plants after transformation with FLARE-S plastid vectors and selection for resistance to spectinomycin and streptomycin, respectively.

Plastid transformation vectors contain a selectable marker gene and passenger gene(s) flanked by homologous plastid targeting sequences (Zoubenko et al., 1994), and are introduced into plastids by biolistic DNA delivery (Svab et al., 1990; Svab and Maliga, 1993) or PEG

treatment (Golds et al., 1993; Koop et al., 1996; O'Neill et al., 1993). The selectable marker genes may encode resistance to spectinomycin, streptomycin or kanamycin. Resistance to the drugs is conferred by the expression of chimeric *aadA* (Svab and Maliga, 1993) and *neo* (*kan*) (Carrer et al., 1993) genes in plastids. These drugs inhibit chlorophyll accumulation and shoot formation on plant regeneration media. The transplastomic lines are identified by the ability to form green shoots on bleached wild-type leaf sections. Obtaining a genetically stable transplastomic line involves cultivation of the cells on a selective medium, during which the cells divide at least 16 to 17 times (Moll et al., 1990). During this time wild type and transformed plastids and plastid genome copies gradually sort out. The extended period of genome and organellar sorting yields chimeric plants consisting of sectors of wild-type and transgenic cells (Maliga, 1993). In the chimeric tissue antibiotic resistance conferred by *aadA* or *neo* is not cell autonomous: transplastomic and wild-type sectors are both green due to phenotypic masking by the transgenic tissue. Chimerism necessitates a second cycle of plant regeneration on a selective medium. In the absence of a visual marker this is an inefficient process, involving antibiotic selection and identification of transplastomic plants by PCR or Southern probing. The feasibility of visual identification of transformed sectors greatly reduces the effort required to obtain homoplastomic clones.

The *Aequorea victoria* green fluorescent protein (GFP) is a visual marker, allowing direct imaging of the fluorescent gene product in living cells without the need for prolonged and lethal histochemical staining

procedures. Its chromophore forms autocatalytically in the presence of oxygen and fluoresces green when absorbing blue or UV light (Prasher et al., 1992; Chalfie et al., 1994; Heim et al., 1994) (reviewed in 5 ref. Prasher, 1995; Cubitt et al., 1995; Misteli and Spector, 1997). The *gfp* gene was modified for expression in the plant nucleus by removing a cryptic intron, introducing mutations to enhance brightness and to improve GFP solubility (Pang et al., 1996; Reichel et 10 al., 1996; Rouwendal et al., 1997; Haseloff et al., 1997; Davis and Vierstra, 1998). GFP was used to monitor protein targeting to nucleus, cytoplasm and plastids from nuclear genes (Sheen et al., 1995; Chiu et al., 1996; Kähler et al., 1997), and to follow virus movement 15 in plants (Baulcombe et al., 1995; Epel et al., 1996). GFP has also been used to detect transient gene expression in plastids (Hibberd et al., 1998).

The expression of GFP by directly incorporating the *gfp* gene in the plastid genome is described herein. 20 Incorporation of a visual marker, the GFP protein, in the plastid transformation vectors of the present invention facilitates distinction of spontaneous antibiotic resistant mutants and plastid transformants (Svab et al., 1990). Furthermore, transplastomic sectors 25 in the chimeric tissue can be visually identified, significantly reducing the time and effort required for obtaining genetically stable transplastomic lines. The utility of the GFP marker described here is further enhanced by its fusion with the enzyme aminoglycoside 30 3''-adenylyltransferase [AAD] conferring spectinomycin and streptomycin resistance to plants. Using a marker gene encoding a bifunctional protein, FLARE-S (fluorescent antibiotic resistance enzyme, spectinomycin

and streptomycin), prevents physical separation of the two genes and simplifies engineering. Furthermore, fluorescent antibiotic resistance genes enables extension of plastid transformation to cereal crops, in which plastid transformation is not associated with a readily identifiable tissue culture phenotype.

The following protocols are provided to facilitate the practice of the present example.

**Construction of tobacco plastid vectors.** The *aadA16gfp* gene encodes FLARE16-S fusion protein, and can be excised as an *NheI*-*XbaI* fragment from plasmid pMSK51, a pBSKSII+ derivative (Genbank Accesssion No. Not yet assigned). The fusion protein was obtained by cloning *gfp* (from plasmid pCD3-326F) downstream of *aadA* (in plasmid pMSK38), digesting the resulting plasmid with *BstXI* (at the 3' end of the *aadA* coding region) and *NcoI* (including the *gfp* translation initiation codon) and linking the two coding regions by a *BstXI*-*NcoI* compatible adapter. The adapter was obtained by annealing oligonucleotides 5'-GTGGGCAAAGAACTTGTTGAA GGAAAATTGGAGCTAGTAGAAGGTCTTAAAGTCGC-3' and 5'-CATGGCGACTTTAAGACCTTCTACTAGCTCCAATTTTCCTTCAACAAGTTCTTTGCCACTACC-3'. The adapter connects AAD and GFP with a peptide of 16 amino acid residues (ELVEGKLELVEGLKVA).

The engineered *aadA* gene (Chinault et al., 1986) in plasmid pMSK38 (pBSIIKS+ derivative) has *NcoI* and *NheI* sites at the 5' end and *BstXI* and *XbaI* sites at the 3' end of the gene. The *NcoI* site includes the translation initiation codon; the *NheI* and *BstXI* sites are in the coding region close to the 5' and 3' ends, respectively; the *XbaI* site is downstream of stop codon. The mutations were introduced by PCR using

oligonucleotides 5'-

GGCCATGGGGGCTAGCGAAGCGGTGATCGCCGAAGTATCG-3' and 5'-  
CGAATTCTAGACATTATTTGCCCACTACCTTGGTGATCTC-3'.

The *gfp* gene in plasmid CD3-326F is the  
5 derivative of plasmid psmGFP, encoding the soluble  
modified version of GFP (accession number U70495)  
obtained under order number CD3-326 from the Arabidopsis  
Biological Resource Center, Columbus, OH (Davis and  
Vierstra, 1998). The *gfp* gene in plasmid CD3-326F is  
10 expressed in the PpsbA /TpsbA expression cassette. The  
*gfp* gene in plasmid CD3-326F was obtained through the  
following steps. The BamHI-SacI fragment from CD3-326  
was cloned into pBSKS+ vector to yield plasmid CD3-326A.  
The SacI site downstream of the coding region was  
15 converted into an XbaI site by blunting and linker  
ligation (5'-GCTCTAGAGC; plasmid CD3-326B). An NcoI site  
was created to include the translation initiation codon  
and at the same time the internal NcoI site was removed  
by PCR amplification of the coding region N-terminus  
20 with primers 5'-  
CCGGATCCAAGGAGATATAACACCATGGCTAGTAAAGGAGAAGAAGTTTTC-3'  
and 5'-GTGTTGGCCAAGGAACAGGTAGTTTTCC-3'. The PCR-  
amplified fragment was digested with BamHI and MscI  
restriction enzymes, and the resulting fragment was used  
25 to replace the BamHI-MscI fragment in plasmid CD3-326B  
to yield plasmid CD3-326C. The *gfp* coding region was  
excised from plasmid CD3-326C as an NcoI-XbaI fragment  
and cloned into a *psbA* cassette to yield plasmid CD3-  
326D. PpsbA and TpsbA are the *psbA* gene promoter and  
30 3'- untranslated region derived from plasmids pJS25  
(Staub and Maliga, 1993). TpsbA has been truncated by  
inserting a HindIII linker downstream of the modified  
BspHI site (Peter Hajdukiewicz, unpublished). The



PpsbA::gfp::TpsbA gene was excised as an EcoRI-HindIII fragment and cloned into EcoRI and HindIII digested pPRV111A, to yield plasmid CD3-326F.

5 The *aadA16gfp* coding region from plasmid pMSK51 was introduced into two expression cassettes. In plasmid pMSK53 the *aadA16gfp* coding region is expressed in the PrnLrbcL+DBwt/TpsbA cassette, and encodes the FLARE16-S2 protein (fluorescent antibiotic resistance enzyme, spectinomycin). PrnLrbcL+DBwt is described in the  
10 previous examples and derives from plasmid pHK14. The construct contains a chimeric promoter composed of the *rrn* operon promoter, the *rbcL* gene leader and downstream box sequence. TpsbA is the *psbA* gene 3' untranslated region, and functions to stabilize the chimeric mRNA. In  
15 plasmid pMSK54 the *aadA16gfp* coding region is expressed in the PrnLatpB+DBwt/TpsbA cassette, and encodes the FLARE16-S1 protein. PrnLatpB+DBwt derives from plasmid pHK10, and is a chimeric promoter composed of the *rrn* operon promoter, the *atpB* leader and downstream box  
20 sequence. See Examples 1-4.

The chimeric *aadA16gfp* genes were introduced into the tobacco plastid transformation vector pPRV111B (Zoubenko et al., 1994). The *aadA* gene was excised from  
25 plasmid pPRV111B with EcoRI and SpeI restriction enzymes, and replaced with the EcoRI-SpeI fragment from plasmids pMSK53 and pMSK54 to generate plasmids pMSK57 (*aadA16gfp-S2*) and pMSK56 (*aadA16gfp-S1*).

**Construction of rice plastid vectors.** Plasmid  
30 pMSK49 is a rice-specific plastid transformation vector which carries the *aadA11gfp-S3* gene as the selective marker in the *trnV/rps12/7* intergenic region (GenBank Accession Number: Not yet assigned). Plasmid pMSK49

carries the rice *Sma*I-*Sna*BI plastid fragment (restriction sites at nucleotides 122488 and 125 878 in the genome Hiratsuka et al., 1989) cloned into a pBSKSII+ (Stratagene) vector after blunting the *Sac*I and  
5 *Kpn*I restriction sites. The *Xba*I site present in the rice plastid DNA fragment (position at nucleotide 125032 in the genome (Hiratsuka et al., 1989) was removed by filling in and religation. Prior to cloning the selective marker the progenitor plasmid was digested  
10 with the *Bgl*II restriction enzyme giving rise to a deletion of 119 nucleotides between two proximal *Bgl*II sites (positions at 124367 and 124491). The *aadA11gfp-S3* gene was then cloned in the blunted *Bgl*II sites.

The *aadA* gene in plasmid pMSK49 was obtained by  
15 modifying the *aadA* gene in plasmid pMSK38 (above) to obtain plasmid pMSK39. The modification involved translationally fusing the *aadA* gene product at its N-terminus with an epitope of the human c-Myc protein (amino acids 410-419; EQKLISEEDL Kolodziej and Young,  
20 1991). The genetic engineering was performed by ligating an adapter obtained by annealing complementary oligonucleotides with appropriate overhangs into *Nco*I-*Nhe*I digested pMSK38 plasmid. The oligonucleotides were:  
25 5'-CATGGGGGCTAGCGAACAAAACTCATTTCTGAAGAAGACTTGc-3' and 5'-CTAGGCAAGTCTTCTTCAGAAATGAGTTTTTGTTCGCTAGCCCC-3'.

The *aadA11gfp* gene encoding FLARE11-S was obtained by linking AAD and GFP with the 11-mer peptide ELAVEGKLEVA. To clone *aadA* and *gfp* in the same polycloning site, *gfp* (*Eco*RI-*Hind*III fragment; from  
30 plasmid CD3-326F) was cloned downstream of *aadA* in plasmid pMSK39 to obtain plasmid pMSK41. The two genes were excised together as an *Nhe*I-*Hind*III fragment, and cloned into plasmid pMSK45 to replace a kanamycin-

resistance gene yielding plasmid pMSK48. Plasmid pMSK45 is a derivative of plasmid pMSK35 which carries the PrnLT7g10+DB/Ec promoter. The promoter consists of the plastid rRNA operon promoter and the leader sequence of the T7 phage gene 10 leader. In plasmid pMSK48, *aadA* is expressed from the PrnLT7g10+DB/Ec promoter. The *aadA* and *gfp* genes were then translationally fused with an BstXI-NcoI adapter that links the AAD and GFP with an 11-mer peptide. The adapter was obtained by annealing oligonucleotides 5'-  
GTGGGCAAAGAACTTGCAGTTGAAGGAAAATTGGAGGTCGC-3' and 5'-  
CATGGCGACCTCCAATTTTCCTTCAACTGCAAGTTCTTTGCCCACTACC-3',  
which was ligated into BstXI/NcoI digested pMSK48 plasmid DNA to yield plasmid pMSK49. Plasmid pMSK49 has the rice plastid targeting sequences present in plasmid pMSK35.

**Tobacco plastid transformation.** Tobacco leaves from 4 to 6 weeks old plants were bombarded with DNA-coated tungsten particles using the Dupont PDS1000He Biolistic gun (1100 psi). Transplastomic clones were identified as green shoots regenerating on bleached leaf sections on RMOP medium containing 500mg/L spectinomycin dihydrochloride (Svab and Maliga, 1993). The spectinomycin resistant shoots were illuminated with UV light (Model B 100AP, UV Products, Upland, California, USA). Shoots emitting green light were transferred to spectinomycin free MS medium (Murashige and Skoog, 1962) (3% sucrose) on which fluorescent (transplastomic) and non-fluorescent (wild-type) sectors formed. Fluorescent sectors were excised, and transferred to selective (500 mg/L spectinomycin) shoot regeneration (RMOP) medium. Regenerated shoots were tested for uniform transformation by Southern analysis.

**Rice plastid transformation.** Callus formation from mature *Oryza sativa* cv. Taipei 309 seeds was induced on a modified CIM medium (Tompson et al., 1986), containing MS salts and vitamins (2 mg/L glycine, 0.5 mg/L nicotinic acid, 0.5 mg/L pyridoxine and 0.1 mg/L thiamine), 2 mg/L 2,4D, 1 mg/L kinetin and 300 mg/L casein enzymatic hydrolysate Type III (Sigma C-1026) and sucrose (30g/L). Embryogenic suspensions from the proliferating embryogenic calli were obtained on the AA medium (Muller and Grafe, 1978). For plastid transformation by the biolistic process rice embryogenic cells were plated on a filter paper on non-selective modified CIM medium (Tompson et al., 1986). The bombarded cells were incubated for 48 hours, transferred to selective liquid AA medium (Muller and Grafe, 1978) (one to two weeks), and then to solid modified RRM regeneration medium (Zhang and Wu, 1988) containing MS salts and vitamins, 100 mg/L myo-inositol, 4 mg/L BAP, 0.5 mg/L IAA, 0.5 mg/L NAA, 30 g/L sucrose and 40 g/L maltose and 100 mg/L streptomycin sulfate on which green shoots appeared in two to three weeks. The shoots were rooted on a selective MS salt medium (Murashige and Skoog, 1962) containing 30 g/L sucrose and 100 mg/L streptomycin sulfate. Leaf samples for PCR analysis and confocal microscopy were taken from plants on selective medium.

**PCR amplification of border fragments.** Total cellular DNA was extracted according to Mettler (Mettler, 1987). The PCR analysis was carried out with a 9:1 mixture of AmpliTaq (Stratagene) and Vent (New England Biolabs) DNA polymerases in the Vent buffer following the manufacturer's recommendations. The left

border fragment was amplified with primers O3 (5'-ATGGATGAACTATACAAATAAG-3') and O4 (5'-GCTCCTATAGTGTGACG-3'). The right border fragment was amplified with primers O5 (5'-ACTACCTCTGATAGTTGAGTCG-3') and O6 (5'-AGAGGTTAATCGTACTCTGG-3'). The aadA part of FLARE-S genes was amplified with primers O1 (5'-GGCTCCGCAGTGGATGGCGGCCTG-3') and O2 (5'-GGGCTGATACTGGGCCGCGCAGG-3'). Primer positions are shown in Fig. 5A. Note that the same primers can be used in transplastomic tobacco and rice plants expressing FLARE-S.

**Detection of FLARE-S by fluorescence.** FLARE-S expressing sectors in the leaves were visualized by an Olympus SZX stereo microscope equipped for GFP detection with a CCD camera system. Subcellular localization of GFP was verified by laser-scanning confocal microscopy (Sarastro 2000 Confocal Image System, Molecular Dynamics, Sunnyvale, CA). This system includes an argon mixed gas laser with lines at 488 and 568 nm and detector channels. The channels are adjusted for fluorescein and rhodamine images. GFP fluorescence was detected in the FITC channel (488-514 nm). Chlorophyll fluorescence was detected in the TRITC channel (560-580 nm). The images produced by GFP and chlorophyll fluorescence were viewed on a computer screen attached to the microscope and processed using the Adobe PhotoShop software.

**Immunoblot analysis.** Leaves (0.5 g) collected from plants in sterile culture were frozen in liquid nitrogen and ground to a fine powder in a mortar with a pestle. For protein extraction the powder was transferred to a

centrifuge tube containing 1 ml buffer [50 mM Hepes/KOH (pH 7.5), 1 mM EDTA, 10 mM potassium acetate, 5 mM magnesium acetate, 1 mM dithiothreitol and 2 mM PMSF] and mixed by flicking. The insoluble material was removed by centrifugation at 4°C for 5 min at 11,600 *g*. Protein concentration in the supernatant was determined using the Biorad protein assay reagent kit. Proteins (20 µl per lane) were separated in 12% SDS-PAGE (Laemmli, 1970). Proteins separated by SDS-PAGE were transferred to a Protran nitrocellulose membrane (Schleicher and Schuell) using a semi-dry electroblotting apparatus (Bio-Rad). The membrane was incubated with Living Colors Peptide Antibody (Clontech) diluted 1 to 200. FLARE-S was visualized using ECL chemilluminescence immunoblot detection on X-ray film. FLARE-S on the blots was quantified by comparison with a dilution series of commercially available purified wild-type GFP (Clontech).

#### RESULTS AND DISCUSSION

Tobacco plastid vectors with FLARE-S as the selectable marker.

Two FLARE-S fusion proteins were tested in *E. coli*. In one, the AAD and GFP were linked by an 11-mer (ELAVEGKLEVA), in the second by a 16-mer (ELVEGKLELVEGLKVA) linker. For transformation in tobacco, the *aadA16gfp* coding region (16-mer linker) was expressed in two cassettes known to mediate high levels of protein accumulation in plastids. Both utilize the strongest known plastid promoter driving the expression of the ribosomal RNA operon (*Prrn*), and the 3'-UTR of the highly expressed *psbA* gene (*TpsbA*) for the stabilization of the chimeric mRNAs. The *PrrnLatpB+wtDB* (plasmid pMSK56) and *PrrnLrbcL+DBwt* (plasmid pMSK57)

promoters utilize the *atpB* or *rbcL* gene leader sequences and the coding region N-termini with the downstream box (DB) sequence, respectively. Due to inclusion of the DB sequence in the chimeric genes, the proteins encoded by the two genes are slightly different, having 14 amino acids of the ATP-ase  $\beta$  subunit (*atpB* gene products) or ribulose 1,5-bisphosphate carboxylase/oxygenase (*rbcL* gene product) translationally fused with FLARE16-S (FLARE16-S1 and FLARE16-S2, respectively). To obtain a plastid transformation vector with the fluorescent spectinomycin resistance genes, the chimeric genes were cloned into the *trnV/rps12/7* plastid intergenic region in plastid vector pPRV111B. Plasmids pMSK56 and pMSK57 (Fig. 23) express FLARE16-S1 and FLARE16-S2, respectively, as markers.

**Identification of transplastomic tobacco clones by fluorescence.** Transformation was carried out by biolistic delivery of pMSK56 and pMSK57 plasmid DNA into chloroplast. The bombarded leaves were transferred onto selective (500 mg/L spectinomycin) shoot regeneration medium. Wild-type leaves on this medium bleach and form white callus. Cells with transformed plastids regenerate green shoots. The leaves on the selective medium were regularly inspected with a hand-held long-wave UV lamp for FLARE-S fluorescence.

No fluorescence could be detected in young shoots (3 to 5 mm in size) developing on pMSK56-bombarded leaves. However, formation of bright sectors in the leaves was observed, when these small shoots were transferred onto non-selective plant maintenance medium. In contrast, cultures bombarded with plasmid pMSK57 yielded small fluorescent shoots at an early stage.

These fluorescent shoots, and some of the non-fluorescent ones, developed into plants with bright sectors on non-selective plant maintenance medium. Therefore, FLARE16-S2 is useful for early detection of plastid transformation events. FLARE16-S2 fluorescence in young shoots on a selective medium should be due to relatively high levels of FLARE16-S2. Higher levels of FLARE16-S2 are also indicated by the brighter sectors in variegated leaves expressing FLARE16-S2 as compared to FLARE16-S1.

The size of sectors was different in individual shoots. FLARE-S expression in different leaf layers was also obvious. With the traditional selection for spectinomycin resistance, the transplastomic and wild-type sectors are not visible. Regeneration of plants with uniformly transformed plastid genomes was greatly facilitated by the fluorescing sectors expressing FLARE-S, which could be readily identified in UV light, dissected, and transferred for a second cycle of plant regeneration on spectinomycin-containing (500 mg/L) selective medium.

Given the high levels of FLARE-S accumulation we were interested to find out, if FLARE-S is toxic to plants. We expected that toxicity should be manifested as lower transformation efficiencies. Bombardment of 30 tobacco leaves with plasmids pMSK56 and pMSK57 yielded 71 and 89 spectinomycin resistant clones, respectively. Out of these, 61 and 77 lines were verified as transplastomic by fluorescence. Plastid transformation in a subset of these was confirmed by confocal laser scanning microscopy (7 clones each; see below) and Southern analysis (4 clones). The frequency of plastid transformation events with the FLARE-S -expressing genes



was slightly higher (~2 instead of ~1 per bombardment) than reported earlier with a chimeric *aadA* gene at the same insertion site (Svab and Maliga, 1993). Therefore, we assume that accumulation of FLARE-S at high levels is not detrimental. Lack of toxicity is also supported by the apparently normal phenotype of the plants in the greenhouse (not shown).

**Localization of FLARE-S to tobacco plastids by confocal microscopy.** Due to phenotypic masking, transplastomic and wild type sectors in a chimeric leaf are both green on a selective medium. However, we have found that in chimeric leaf sectors in the same cell some plastids express FLARE-S while others do not, when observed by confocal microscopy (Fig. 24). FLARE-S and chlorophyll fluorescence were detected separately in the fluorescein and rhodamine channels, respectively. The two images were then overlaid confirming that FLARE-S fluorescence derives from chloroplasts.

Expression of FLARE-S was also studied in non-green plastid types including the chromoplasts in petals and the non-green plastids in root cells (Fig. 24b,f). These studies were carried out in plants, which were homoplastomic for the transgenomes. Homoplastomic state was important, since in non-green tissues chlorophyll could not be used for confirmation of the organelles as plastids. Since FLARE-S expression could be readily detected in chloroplasts as well as non-green plastids, the plastid rRNA operon promoter is apparently active in all plastid types.

#### **FLARE-S accumulation in tobacco leaves.**

Accumulation of FLARE-S in homoplastomic leaves was

tested using the commercially available GFP antibody, recognizing the GFP portion (239 amino acid residues) of FLARE16-S (520 amino acids). FLARE16-S1 (532 amino acids) was ~8 %, whereas FLARE16-S2 (532 amino acids) was ~18 % of total soluble leaf protein (Fig. 25). To calculate FLARE16-S concentrations, a GFP dilution series was used as a reference, and the values were then increased by 2.6 to correct for the larger size of the FLARE16-S1 and -S2 proteins.

**Tracking plastid transformation in rice by FLARE-S expression.**

In rice, plant regeneration is from non-green embryogenic cells. Encouraged by FLARE-S expression in non-green tobacco plastids, we attempted to transform the non-green plastids of embryogenic rice tissue-culture cells. Plastid transformation was carried out using a rice-specific vector expressing FLARE11-S3 and targeting insertion of the *aadA11gfp-S3* gene in the *trnV/rps12/7* intergenic region. The location of the insertion site and the size of plastid targeting sequences in the rice vector are similar to the tobacco vectors shown in Fig. 23.

Plastid transformation in rice was carried out by bombardment of embryogenic rice suspension culture cells using gold particles coated with plasmid pMSK49 DNA. Rice cells, as most cereals, are naturally resistant to spectinomycin (Fromm et al., 1987). FLARE-S, however, confers resistance to streptomycin as well (Svab and Maliga, 1993). Therefore, selection for transplastomic lines was carried out on selective streptomycin medium (100 mg/L). Streptomycin at this concentration inhibits the growth of embryogenic rice cells. After bombardment, the rice cells were first

selected in liquid embryogenic AA medium, then on the solid plant regeneration medium, on which the surviving resistant cells regenerated green shoots (12 in 25 bombarded plates). These shoots were rooted, and grown  
5 into plants. PCR amplification of border fragments in DNA isolated from the leaves of these plants confirmed integration of *aadA11gfp-S3* sequences in the plastid genome (Fig. 26). The left and right border fragments can not be amplified if the gene is integrated into the  
10 nuclear genome, as one of the primers (O4 or O6) of the pairs is outside the plastid targeting regions.

FLARE11-S3 expression in the leaves of two of the PCR-positive plants was tested by confocal laser-scanning microscopy. In rice, as in tobacco, the FLARE-S  
15 marker confirmed segregation of transplastomic and wild-type plastids (Fig. 27). In rice only a small fraction of chloroplasts expressed FLARE-S. Since individual cells marked with arrows in Fig. 27 contained a mixed population of wild-type and transgenic chloroplasts,  
20 FLARE-S in these cells could be expressed only from the plastid genome. Integration of *aadA11gfp-S3* into the nuclear genome downstream of plastid-targeting transit peptide would result in uniform expression of FLARE-S in each of the chloroplasts within the cell.

25 The sequences of the selectable marker genes of the invention are provided in Figures 28-34. Figure 35 depicts a table describing the selectable marker genes disclosed in the present example.

30 Direct visual identification of transplastomic sectors requires high level expression of FLARE-S in plastids. High GFP expression levels in Arabidopsis were toxic, interfering with plant regeneration. Toxicity of

wild-type (insoluble) GFP was linked to GFP accumulation in the nucleus and cytoplasm, and could be eliminated by targeting it to the endoplasmic reticulum (Haseloff et al., 1997). GFP aggregates were also cytotoxic to *E. coli* cells (Cramer et al., 1996). To enhance fluorescence intensity and to avoid cytotoxicity, soluble versions of the codon-modified GFP were obtained (Davis and Vierstra, 1998). We have utilized the gene for a soluble-modified GFP described by Davis and Vierstra (Davis and Vierstra, 1998) to create variants of FLARE-S, a fusion protein, which does not have an apparent cytotoxic effect. The frequency of plastid transformation, if affected at all, is increased rather than decreased. In tobacco, we normally obtain one transplastomic clone per bombarded leaf sample (Svab and Maliga, 1993), whereas with the FLARE-S genes on average we could recover two clones per sample. Plant regeneration from highly fluorescent tissue was readily obtained, and the regenerated plants have a phenotype indistinguishable from the wild type.

Plastid transformation in rice requires expression of the selective marker in non-green plastids. The rRNA operon has two promoters, one for the eubacterial-type (PEP) and one for the phage-type (NEP) plastid RNA polymerase. The promoter driving FLARE-S expression is recognized only by the eubacterial-type plastid RNA polymerase. Previously, it was assumed that the eubacterial-type promoter is active only in chloroplasts (Maliga, 1998). Accumulation of FLARE-S in roots and petals indicates that PEP is also active in non-green plastids.

Plastid transformation is a process that unavoidably yields chimeric plants, since cells of

higher plants contain a large number (300 to 50000) of plastid genome copies (Bendich, 1987), out of which initially only a few are transformed. High level expression of FLARE-S in plastids provides the means for visual identification of transplastomic sectors, even if they are present in a chimeric tissue. GFP and AAD could be expressed from two different genes in a plastid transformation vector. However, transformation with a marker gene encoding a bifunctional protein prevents separation of the two genes and simplifies engineering. The fluorescent selective marker will significantly reduce the work required to obtain genetically stable plastid transformants in tobacco, a species in which plastid transformation is routine. The bottleneck of applying plastid transformation in crop improvement is the lack of technology. In tobacco, chimeric clones with transformed plastids are readily identified by shoot regeneration (Svab et al., 1990). In Arabidopsis, clones with transformed plastids are identified by greening (Sikdar et al., 1998). We have shown here that FLARE-S is a suitable marker to select for transplastomes in embryogenic rice cells, which lack the visually identifiable tissue culture phenotypes exploited in tobacco and Arabidopsis. Data presented here are the first example for stable integration of foreign DNA into the rice plastid genome. These rice plants are heteroplastomic. Uniformly transformed rice plants will be obtained by further selection on streptomycin medium and screening the embryogenic cells for FLARE-S expression. Thus, the FLARE-S marker system will enable extension of plastid transformation to cereal crops.

#### **The utility of the new chimeric promoters**

The  $\sigma^{70}$ -type plastid ribosomal RNA operon promoter, Prn, is the strongest known plastid promoter expressed in all tissue types. The ultimate product of this promoter in the plastid is RNA not protein. Therefore, a series of chimeric promoters were constructed to facilitate protein accumulation from Prn, using expression of the neomycin phosphotransferase (NPTII) enzyme as the reference protein.

10

1) The expression cassettes have distinct tissue-specific expression profiles. Some of the expression cassettes described here will facilitate relatively high levels of protein expression in all tissues, including leaves, roots and seeds. Other cassettes have different expression profiles: for example will facilitate moderate levels of protein accumulation in the leaves while lead to relatively high levels of protein accumulation in the roots. Accumulation of a protein at levels of 10% to 50% of total soluble protein is considered high-level protein expression; low-levels of protein expression would be in the range of  $\leq 0.1\%$  total soluble cellular protein.

15

20

25

30

2) Efficiency of the selectable marker gene depends on the rate at which the gene product accumulates during the early stage of transformation. Since initially present only in a few copies per cell, high levels of expression from a few copies will provide protection from toxic substances early on, facilitating efficient recovery of transformed lines. The expression cassettes will be useful to drive the expression of the genes conferring resistance to the antibiotics

streptomycin, spectinomycin and hygromycin, and the herbicides phosphinotrycin and glyphosate. In such applications addition of amino acids at the N-terminus is acceptable, as long as it does not interfere with the expression of the selectable marker genes. NPTII is such an enzyme. In cases like NPTII, an N-terminal fusion and thereby the mRNA "Downstream Box" sequences give an additional at least two to four-fold increase in protein levels. The -DB construct which relied on an *NheI* site, and involved addition of one (N-terminal) amino acid of the source gene coding region is convenient, but is not necessary. When translational fusion is not feasible due to inactivation of proteins, seamless in-frame constructs may be created by PCR methods outlined in the application.

3) A second major area on which application of the chimeric promoters is extremely useful is protein expression for pharmaceutical, industrial or agronomic purposes. The examples include, but are not restricted to, production of vaccines, healthcare products like human hemoglobin, industrial or household enzymes.

#### REFERENCES

- Allen, G.C., Hall, G.J., Michalowski, S., Newman, W., Spiker, S., Weissinger, A.K. and Thompson, W.F. (1996) High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. *Plant Cell*, 8, 899-913.
- Allison LA, Simon LD, Maliga P (1996) Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. *EMBO J* 15: 2802-2809
- Arntzen, C.J. (1997) High-tech herbal medicine: plant-based vaccines [news]. *Nature Biotechnology*, 15, 221-222.

- Baulcombe, D. C. Chapman, S. and Cruz, S. S. 1995. Jellyfish green fluorescent protein as a reporter for virus infections. *Plant J.* 7: 1045-1053.
- 5 Beck, E., Ludwig, G., Auerswald, E.A., Reiss, B. and Schaller, H. (1982) Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene*, 19, 327-336.
- 10 Bendich, A. J. 1987. Why do chloroplasts and mitochondria contain so many copies of their genome? *Bio-essays* 6: 279-282.
- Bonham-Smith, P.C. and Bourque, D.P. (1989) Translation of chloroplast-encoded mRNA: potential initiation and termination signals. *Nucleic Acids Res.* 17, 2057-2080.
- 15 Cao, J., Duan, X., McElroy, D., and Wu, R. (1992) Regeneration of herbicide resistant transgenic rice plants following microprojectile-mediated transformation of suspension culture cells. *Plant Cell Report* 11:586-591.
- 20 Carrer, H. and Maliga, P. (1995) Targeted insertion of foreign genes into the tobacco plastid genome without physical linkage to the selectable marker gene. *Biotechnology*, 13, 791-794.
- 25 Carrer, H., Hockenberry, T.N., Svab, Z. and Maliga, P. (1993) Kanamycin resistance as a selectable marker for plastid transformation in tobacco. *Mol. Gen. Genet.*, 241, 49-56.
- 30 Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263: 802-805.
- Chaudhuri, S. and Maliga, P. (1996) Sequences directing C to U editing of the plastid psbL mRNA are located within a 22 nucleotide segment spanning the editing site. *EMBO J.*, 15, 5958-5964.
- 35 Chinault, A.C., Blakesley, V.A., Roessler, E., Willis, D.G., Smith, C.A., Cook, R.G., and Fenwick, R.G. 1986. Characterization of transferable plasmids for *Shigella flexneri* 2a that confer resistance trimethoprim, streptomycin and sulfonamides. *Plasmid* 15: 119-131.
- 40 Chiu, W-L., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J. 1996. Engineered gfp as a vital reporter in plants. *Curr. Biol.* 6: 325-330.
- 45 Conrad, U. and Fiedler, U. (1998) Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. *Plant Mol. Biol.*, 38, 101-109.



- Corriveau, J.L., and Coleman, A.W. (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and the results for over 200 angiosperm species. *Amer. J. Bot.* 75:1443-1458.
- 5 Cramer, A., Whitehorn, E. A., Tate, E. and Stemmer, W. P. C. 1996. Improved green fluorescent protein by molecular evolution by DNA shuffling. *Nature Biotech.* 14: 315-319.
- 10 Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A. and Tsien, R. Y. 1995. Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* 20: 448-455.
- Dams, E., Hendriks, L., Van de Peer, Y., Neefs, J.M., Smits, G., Vandenbempt, I. and De Wachter, R. (1988) Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res.*, 16 Suppl, r87-173.
- 15 Daniell, H., Datta, R., Varma, S., Gray, S., and Lee, S.B. (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nat. Biotech.* 16:345-348.
- 20 Davis, S.J., and Vierstra, R.D. 1998. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* 36: 521-528.
- 25 De Block, M., Botterman, J., Vandewiele, M., Dockx, J., Thoen, C., Gossele, V., Rao, Movva, N., Thompson, C., Van Montagu, M., and Leemans, J. (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *EMBO J.* 6:2513-2518.
- 30 De Block, M., De Brouwer, D., Tenning, P. (1989). Transformation of *Brassica napus* and *Brassica oleracea* using *Agrobacterium tumefaciens* and the expression of the bar and neo genes in the transgenic plants. *Plant Physiol.* 91:694-701.
- 35 Deana, A., Ehrlich, R. and Reiss, C. (1998) Silent mutations in the *Escherichia coli* ompA leader peptide region strongly affect transcription and translation in vivo. *Nucleic Acids Res.*, 26, 4778-4782.
- 40 Ellis, R.J. (1979) The most abundant protein in the world. *Trends Biochem. Sci.*, 4, 241-244.
- Epel, B.L., Padgett, H.S., Heinlein, M., and Beachy, R. 1996. Plant virus movement protein dynamics probed with GFP-protein fusion. *Gene* 173: 75-79.
- 45 Etchegaray, J.P. and Inouye, M. (1999) Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. *J. Biol. Chem.*, 274, 10079-10085.
- Faxin, M., Plumbridge, J. and Isaksson, L.A. (1991) Codon choice and potential complementarity between mRNA

downstream of the initiation codon and bases 1471-1480 in 16S ribosomal RNA affects expression of glnS. *Nucleic Acids Res.*, 19, 5247-5251.

- 5 Fromm, H., Edelman, M., Aviv, D. and Galun, E.  
1987. The molecular basis of rRNA-dependent spectinomycin resistance in *Nicotiana* chloroplasts. *EMBO J.* 11: 3233-3237.

- 10 Gallo-Meagher, M., and Irvine, J.E. (1996)  
Herbicide resistant transgenic sugarcane plants containing the bar gene. *Crop Sci.* 36:1367-1374.

Golds, T., Maliga, P., and Koop, H.U. 1993. Stable plastid transformation in PEG-treated protoplasts of *Nicotiana tabacum*. *Biotechnology* 11: 95-97.

- 15 Gray, A.J., and Raybould, A.F. (1988) Reducing  
transgene escape routes. *Nature* 392:653-654.

Hajdukiewicz, P., Allison, L.A., Maliga, P. (1997)  
The two plastid RNA polymerases encoded by the nuclear and plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* 16, 4041-4048.

- 20 Haseloff, J., Siemering, K. R., Prasher, D. C., and  
Hodge, S. 1997. Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proc. Natl. Acad. Sci. USA* 94: 2122-2127.

- 25 Hecker, K.H. and Roux, K.H. (1996) High and low  
annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. *Biotechniques*, 20, 478-485.

- 30 Heim, R., Prasher, D. C. and Tsien, R. Y. 1994.  
Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* 91: 12501-12504.

- 35 Hibberd, J.M., Linley, P.J., Khan, M.S., and Gray,  
J.C. 1998. Transient expression of green fluorescent protein in various plastid types following micro-projectile bombardment. *Plant J.* 16: 627-632.

- 40 Hiratsuka, J., Shimada, H., Whittier, R.,  
Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C-R., Meng, B-Y., Li, U-Q., Kanno, A.,  
Nishizawa, Y., Hirai, A., Shinozaki, K., and Sugiura, M.  
1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217: 185-194.

45 Hirose, T. and Sugiura, M. (1996) Cis-acting elements and trans-acting factors for accurate translation of chloroplast psbA mRNAs: development of an

in vitro translation system from tobacco chloroplasts.  
EMBO J., 15, 1687-1695.

5 Houdt, H.V., Ingelbrecht, I., Montagu, M.V. and  
Depicker, A. (1997) Post-transcriptional silencing of a  
neomycin phosphotransferase II transgene correlates with  
the accumulation of unproductive RNAs and with increased  
cytosine methylation of 3' flanking positions. Plant J.,  
12, 379-392.

10 Ito, K., Kawakami, K. and Nakamura, Y. (1993)  
Multiple control of Escherichia coli lysyl-tRNA  
synthetase expression involves a transcriptional  
repressor and a translational enhancer element. Proc.  
Nat'l. Acad. Sci. USA, 90, 302-306.

15 Kane, J.F. (1995) Effects of rare codon clusters on  
high-level expression of heterologous proteins in  
Escherichia coli. Current Opinion In Biotechnology, 6,  
494-500.

20 Khan, M.S. and Maliga, P. (1999) Fluorescent  
antibiotic resistance marker to track plastid  
transformation in higher plants. Nature Biotechnology,  
in press.

Kling, J. (1996) Could transgenic supercrops one  
day breed superweeds? Science 274:180-181.

25 Kvihler, R.H., Cao, J., Zipfel, W.R., Webb, W.W.,  
and Hanson, M.R. 1997. Exchange of protein molecules  
through connections between higher plant plastids.  
Science 276: 2039-2042.

30 Kolodziej, P. A., and Young, R. A. 1991. Epitope  
tagging and protein surveillance. Methods Enzymol. 194:  
508-519.

35 Koop, H.U., Steinmüller, K., Wagner, H., Rvssler,  
C., Eibl, C., and Sacher, L. 1996. Integration of  
foreign sequences into the tobacco plastome via  
PEG-mediated protoplast transformation. Planta 199:  
193-101.

Kusnadi, A., Nikolov, Z. and Howard, J. (1997)  
Production of recombinant proteins in transgenic plants:  
practical considerations. Biotechnology and  
Bioengineering, 56, 473-484.

40 Laemmli, U. K. 1970. Cleavage of structural  
proteins during the assembly of the head of the  
bacteriophage T4. Nature 227: 680-685.

Lefebvre, B., Formstecher, P. and Lefebvre, P.  
(1995) Improvement of the gene splicing overlap method  
(SOE) method. BioTechniques 19: 186-187

45 Maier, R.M., Neckermann, K., Igloi, G.L. and  
Kvssel, H. (1995) Complete sequence of the maize  
chloroplast genome: gene content, hotspots of divergence

and fine tuning of genetic information by transcript editing. *J. Mol. Biol.*, 251, 614-628.

- 5 Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews*, 60, 512-538.

10 Maliga P (1995). Biolistic transformation of tobacco cells with nuclear drug resistance genes. In Maliga P, Klessig D, Cashmore A, Gruissem W, Varner J (eds), *Methods in Plant Molecular Biology-A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 37-52.

Maliga, P. (1993) Towards plastid transformation in flowering plants. *Trends in Biotechnology* 11:101-107.

- 15 Maliga, P. 1998. Two plastid RNA polymerases of higher plants: an evolving story. *Trends Plant Sci.* 3: 4-6.

20 Mayfield, S.P., Cohen, A., Danon, A. and Yohn, C.B. (1994) Translation of the *psbA* mRNA of *Chlamydomonas reinhardtii* requires a structured RNA element contained within the 5' untranslated region. *J. Cell Biol.*, 127, 1537-1545.

25 McBride, K.E., Scaaf, D.J., Daley, M. and Stalker, D. (1994) Controlled expression of plastid transgenes in plants based on a nuclear-encoded and plastid-targeted T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* 91, 7301-7305.

30 McBride, K.E., Svab, Z., Schaaf, D.J., Hogan, P.S., Stalker, D.M. and Maliga, P. (1995) Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. *Biotechnology (N Y)*, 13, 362-365.

35 Medgyesy, P., Pay, A., and Marton, L. (1986) Transmission of paternal chloroplasts in *Nicotiana*. *Mol. Gen. Genet.* 204:195-198.

Mettler, I.J. 1987. A simple and rapid method for minipreparation of DNA from tissue-cultured plant cells. *Plant Mol. Biol. Rep.* 5: 346-349.

40 Mikkelsen, T.R., Andersen, B., and Jorgensen, R.B. (1996). The risk of crop transgene spread. *Nature* 380:31.

Misteli, T., and Spector, D.L. 1997. Applications of the green fluorescent protein in cell biology and biotechnology. *Nature Biotech.* 15: 961-964.

45 Mitta, M., Fang, L. and Inouye, M. (1997) Deletion analysis of *cspA* of *Escherichia coli*: requirement of the AT-rich UP element for *cspA* transcription and the downstream box in the coding region for its cold shock induction. *Mol Microbiol*, 26, 321-335.

- Moll, B., Posby, L., and Maliga, P. 1990. Streptomycin and lincomycin resistance are selective plastid markers in cultured *Nicotiana* cells. *Mol. Gen. Genet.* 221: 245-250.
- 5 Muller, A.J., and Grafe, R. 1978. Isolation and characterization of cell lines of *Nicotiana tobacum* lacking nitrate reductase. *Mol. Gen. Genet.* 161: 67-76.
- Murashige, T. and Skoog, F. (1962) A revised medium for the growth and bioassay with tobacco tissue culture. *Physiologia Plantarum*, 15, 473-497.
- 10 Nickelsen, J., Fleischmann, M., Boudreau, E., Rahire, M. and Rochaix, J.D. (1999) Identification of cis-acting RNA leader elements required for chloroplast psbD gene expression in *Chlamydomonas*. *The Plant Cell*, 11, 957-970.
- 15 Olins, P.O., Devine, C.S., Rangwala, S.H., Kavka, K.S. (1988) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene* 73: 227-235.
- 20 O'Neill, C., Horvath, G.V., Horvath, E., Dix, P.J., and Medgyesy, P. 1993. Chloroplast transformation in plants: polyethylene glycol (PEG) treatment of protoplasts is an alternative to biolistic delivery system. *Plant J.* 3: 729-738.
- 25 Orozco, E.M., Chen, L.J. and Eilers, R.J. (1990) The divergently transcribed *rbcL* and *atpB* genes of tobacco plastid DNA are separated by nineteen base pairs. *Curr. Genet.*, 17, 65-71.
- 30 Pang, S-Z., DeBoer, D.L., Wan, Y., Ye, G., Layton, J.G., Neher, M.K., Armstrong, C.L., Fry, J.E., Hinchee, M.A.W., and Fromm, M.E. 1996. An improved green fluorescent protein gene as avital marker in plants. *Plant Physiol.* 112: 893-900.
- 35 Prasher, D. C. 1995. Using GFP to see the light. *Trends Genet.* 11: 320-323.
- Prasher, D. C., Eckenrode, V. K., Ward, W. W., Predergast, F. G. and Cormier, M. J. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111: 229-233.
- 40 Reichel, C., Mathur, J., Eckes, P., Langenkemper, K., Koncz, C., Schell, J., Reiss, B., and Maas, C. 1996. Enhanced green fluorescence by the expression of an *Aequorea victoria* green fluorescent protein mutant in mono- and dicotyledonous plant cells. *Proc. Natl. Acad. Sci. USA* 93: 5888-5893.
- 45 Rochaix, J.D. (1996) Post-transcriptional regulation of chloroplast gene expression in

- Chlamydomonas reinhardtii. Plant Mol. Biol., 32, 327-341.
- 5      Rouwendal., G.J.A., Mendes, O., Wolbert, E.J.H., and de Boer, A.D., 1997. Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. Plant Mol. Biol. 33: 989-999.
- 10      Saghai-Marooof, M.A., Scriman, K.M., Jorgensen, R.A., and Allard, R.W. (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA 81: 8014-8018.
- 15      Sakamoto, W., Kindle, K.L. and Stern, D.B. (1993) In vivo analysis of Chlamydomonas chloroplast petD gene expression using stable transformation of beta-glucuronidase translational fusions. Proc. Natl. Acad. Sci. USA, 90, 497-501.
- 20      Sawasaki, T., Seki, M., Anzai, H., Irifune, K., and Morikawa, H. (1994). Stable transformation of Arabidopsis with the bar gene using particle bombardment. Transgenic Res. 3:279-286.
- 25      Serino, G., and Maliga, P. (1997) A negative selection scheme based on the expression of cytosine deaminase in plastids. Plant J. 12:697-701.
- 30      Sheen, J., Hwang, S., Niwa, Y., Kobayashi, H., and Galbraith, D.W. 1995. Green fluorescent protein as a new vital marker in plant cells. Plant J. 8: 777-784.
- 35      Shimada, H., and Sugiura, M. (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res. 19:983-995.
- 40      Shinozaki, K. and Sugiura, M. (1982) The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase / oxygenase. Gene 20: 91-102.
- 45      Shinozaki, K., Deno, H., Wakasugi, T. and Sugiura, M. (1986a) Tobacco chloroplast gene coding for subunit I of proton-translocating ATPase: comparison with the wheat subunit I and E. coli subunit b. Curr. Genet., 10, 421-423.
- 50      Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsabayashi, T., Zaita, N., Chungwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Deno, H., Kamogashira, T., Yamada, K., Kasuda, J., Takaiwa, F., Kato, A., Todoh, N., Shimada, H. and Sugiura, M. (1986b) The complete sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J., 5, 2043-2049.

- Sikdar, S.R., Serino, G., Chaudhuri, S., and Maliga, P. 1998. Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Rep.* 18: 20-24.
- 5        Spencer, T.M., Gordon-Kamm, W.J., Daines, R.J., Start, W.G., and Lemaux, P.G. (1990). Bialaphos selection of stable transformants from maize cell culture. *Theor. Appl. Genet.* 79:625-631.
- 10        Sprengart, M.L., Fuchs, E. and Porter, A.G. (1996) The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. *EMBO J.*, 15, 665-674.
- 15        Sriraman, P., Silhavy, D. and Maliga, P. (1998a) The phage-type PclpP-53 plastid promoter comprises sequences downstream of the transcription initiation site. *Nucleic Acids Res.*, 26, 4874-4879.
- 20        Sriraman, P., Silhavy, D. and Maliga, P. (1998b) Transcription from heterologous rRNA operon promoters in chloroplasts reveals requirement for specific activating factors. *Plant Physiol.*, 117, 1495-1499.
- 25        Sriraman, P., Silhavy, D., Maliga, P. (1998b) Transcription from heterologous ribosomal RNA operon promoters in chloroplasts reveals requirement for specific activating factors. *Plant Physiol.* 117: 1495-1499.
- 30        Staub, J., Maliga, P. (1993) Accumulation of D1 polypeptide in tobacco plastids is regulated via the untranslated region of the psbA mRNA. *EMBO J.* 12:601-606
- Staub, J.M. and Maliga, P. (1994) Translation of psbA mRNA is regulated by light via the 5'-untranslated region in tobacco plastids. *Plant J.*, 6, 547-553.
- 35        Stemmer, W.P., Cramer, A., Ha, K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 164, 49-53.
- 40        Stern, D.B., Higgs, D.C. And Yang, J. (1997) Transcription and translation in chloroplasts. *Trends Plant Sci.* 2, 308-315.
- Stiekema WJ, Heidekamp F, Dirkse WG, van Beckum J, deHaan P, ten Bosch C, Louwerse JD (1988) Molecular cloning and analysis of four potato tuber mRNAs. *Plant Mol Biol* 11: 255-269
- 45        Strauch, E., Wohlleben, W., and Pihler, A. (1988) Cloning of phosphinothricin acetyltransferase gene from *Streptomyces viridochromogenes* T4494 and its expression in *Streptomyces lividans* and *Escherichia coli*. *Gene* 63:65-74.
- Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to

- direct expression of cloned genes. *Methods Enzymol.* 185: 60-89.
- 5 Sugita, M. and Sugiura, M. (1984) Nucleotide sequence and transcription of the gene for the 32,000 dalton thylakoid membrane protein from *Nicotiana tabacum*. *Mol. Gen. Genet.* 195: 308-313.
- 10 Svab, Z., Hajdukiewicz, P., and Maliga, P. 1990. Stable transformation of plastids in higher plants. *Proc. Natl. Acad. Sci. USA* 87: 8526-8530.
- 15 Svab, Z., and Maliga, P. (1993). High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA* 90:913-917.
- 20 Tachibana, K., Watanabe, T., Sekizawa, T., and Takematsu, T. (1986) Action mechanism of bialaphos II Accumulation of ammonia in plants treated with bialaphos. *J. Pest. Sci.* 11:33-37.
- 25 Tanaka, M., Obokata, J., Chunwongse, J., Shinozaki, K., Sugiura, M. (1987) Rapid splicing and stepwise processing of a transcript from the *psbB* operon in tobacco chloroplasts: Determination of the intron sites in *petB* and *petD*. *Mol. Gen. Genet.* 209: 427-431.
- 30 Thompson, C.J., Movva, N.R., Tizard, R., Cramer, R., Davies, J.E., Lauwereys, M., Botterman, J. (1987) Characterization of the herbicide-resistance gene bar from *Streptomyces hygroscopicus*. *EMBO J.* 6:2519-2523.
- 35 Timmons, A.M., Charters, Y.M., Crawford, J.W., Burn, D., Scott, S.E., Subbels, S.J., Wilson, N.J., Robertson, A., O'Brian, E.T., Squire, G.R., and Wilkinson, M.J. (1996) Risks from transgenic crops. *Nature* 380:487.
- 40 Tompson, J.A., Abdullah, R., and Cocking, E.C. 1986. Protoplast culture of rice using media solidified with agarose. *Plant Science* 47: 123-133.
- 45 Uchida, K (1992) Recombination and amplification of multiple portions of genomic DNA by a modified polymerase chain reaction. *Anal. Biochem.* 202: 159-161.
- 50 Varshavsky, A. (1996) The N-end rule: Functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* 93, 121423-12149.
- 55 Vera, A., and Sugiura M (1995). Chloroplast rRNA transcription from structurally different tandem promoters: an additional novel-type promoter. *Curr. Genet.* 27, 280-284.
- 60 Voorma, H.O. (1996) Control of translation in prokaryotes. In: *Translational Control*, Hershey, J.W.B., Methews, M.B. and Sonenberg, N., eds. pp. 759-777, Cold Spring Harbor Laboratory Press, Plainview, NY



Wehrmann, A., Vliet, V.A., Opsomer, C., Bottermann J., and Schulz, A. (1996) The bar and pat gene products make them equally applicable for plant engineers. Nat. Biotechnol. 14:1274-1278.

5        Wohlleben, W., Arnold, W., Broer, I., Hillemann, D., Strauch, E. and Pihler, A. (1988) Nucleotide sequence of the phosphinothricin N-acetyltransferase gene from *Streptomyces T|949* and its expression in *Nicotiana tabacum*. Gene 70:25-37.

10        Wu, C.J. and Janssen, G.R. (1996) Translation of vph mRNA in *Streptomyces lividans* and *Escherichia coli* after removal of the 5' untranslated leader. Mol Microbiol, 22, 339-355.

15        Zerges, W., Girard-Bascou, J. and Rochaix, J.D. (1997) Translation of the chloroplast psbC mRNA is controlled by interactions between its 5' leader and the nuclear loci TBC1 and TBC3 in *Chlamydomonas reinhardtii*. Mol. Cell. Biol., 17, 3440-3448.

20        Zhang, W., and Wu, R. 1988. Efficient regeneration of transgenic plants from rice protoplasts and correctly regulated expression of the foreign gene in the plants. Theor. Appl. Gene. 76: 835-840.

25        Zhou, J., Liu, W.J., Peng, S.W., Sun, X.Y. and Frazer, I. (1999) Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. Journal Of Virology, 73, 4972-4982.

30        Zoubenko, O. V., Allison, L. A., Svab, Z. and Maliga, P. 1994. Efficient targeting of foreign genes into the tobacco plastid genome. Nucleic Acids Res. 22: 3819-3824.

While certain of the preferred embodiments of the present invention have been described and specifically  
35        exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

40

What is claimed is:

- 5                   1. A recombinant DNA construct for expressing  
at least one heterologous protein in the plastids of  
higher plants, said construct comprising a 5' regulatory  
region which includes a promoter element, a leader  
sequence and a downstream box element operably linked to  
10 a coding region of said at least one heterologous  
protein, said chimeric regulatory region enhancing  
translational efficiency of an mRNA molecule encoded by  
said DNA construct.
- 15                   2. A vector comprising the DNA construct of  
claim 1.
3. A recombinant DNA construct as claimed in  
claim 1, said 5' regulatory region being selected from  
20 the group consisting of PrnnLatpB+DBwt, SEQ ID NO:1,  
PrnnLatpB-DB, SEQ ID NO:2, PrnnLatpB+DBm, SEQ ID NO:3,  
PrnnLclpP+DBwt, SEQ ID NO: 4, PrnnclpP-DB, SEQ ID NO:5,  
PrnnLrbcL+DBwt, SEQ ID NO:6, PrnnLrbcL-DB, SEQ ID NO:7,  
PrnnLrbcL+DBm, SEQ ID NO:8, PrnnLpsbB+DBwt, SEQ ID NO:9,  
25 PrnnLpsbB-DB, SEQ ID NO:10, PrnnLpsbA+DBwt, SEQ ID NO:  
11, PrnnLpsbA-DB, SEQ ID NO:12, PrnnLpsbA-DB(+GC), SEQ  
ID NO:13.
- 30                   4. A recombinant DNA construct as claimed in  
claim 1, said 5' regulatory region being selected from  
the group consisting of PrnnLT7g10+DB/Ec, SEQ ID NO:14,

PrrnLT7g10+DB/pt, SEQ ID NO:15, PrrnLT7g10-DB, SEQ ID NO:15.

5 5. A vector comprising a DNA construct as claimed in claim 1.

6. A DNA construct as claimed in claim 1, said downstream box element having a sequence selected from the group consisting of  
10 5'TCCAGTCACTAGCCCTGCCTTCGGCA'3 and  
5'CCCAGTCATGAATCACAAAGTGGTAA'3.

7. A DNA construct as claimed in claim 1, wherein said heterologous protein is expressed from a  
15 bar gene encoded by *S. hydroscopicus* said bar gene inserted into a plasmid selected from the group consisting of pK012, and pJEK3, said pJEK3 having the sequence of SEQ ID NO: 18.

20 8. A DNA construct as claimed in claim 1, wherein said heterologous protein is expressed from a synthetic bar encoding nucleic acid, said synthetic bar nucleic acid having selected from the group consisting of SEQ ID NO: 19 and SEQ ID NO:20.

25 9. A DNA construct as claimed in claim 1, said at least one heterologous protein comprising a fusion protein.

30 10. A DNA construct as claimed in claim 9, said fusion protein having a first and second coding region operably linked to said 5' regulatory region such that production of said fusion protein is regulated by

said 5' regulatory region, said first coding region encoding a selectable marker gene and said second coding region encoding a fluorescent molecule to facilitate visualization of transformed plant cells.

5

11. A vector comprising the DNA construct of claim 10.

12. A DNA construct as claimed in claim 9,  
10 said fusion protein consisting of an aadA coding region operably linked to a green fluorescent protein coding region.

13. A DNA construct as claimed in claim 10,  
15 said aadA coding region being operably linked to said green fluorescent protein coding region via a nucleic acid molecule encoding a peptide linker having a sequence selected from the group consisting of ELVEGKLELVEGLKVA and ELAVEGKLEVA.

20

14. A DNA construct as claimed in claim 10, said construct having a sequence selected from the group of SEQ ID NOS: 21-25 and 27.

25

15. A plasmid for transforming the plastids of higher plants, said plasmid being selected from the group consisting of pHK30(B), pHK31(B), pHK60, pHK32(B), pHK33(B), pHK34(A), pHK35(A), pHK64(A), pHK36(A), pHK37(A), pHK38(A), pHK39(A), pHK40(A),  
30 pHK41(A), pHK42(A), pHK43(A), pMSK56, pMSK57, pMSK48, pMSK49, pMSK35, pMSK53 and pMSK54.

16. A transgenic plant containing a plasmid as claimed in claim 15.

5 17. A transgenic plant as claimed in claim 15, said plant being selected from the group consisting of monocots and dicots.

18. A method for producing transplastomic monocots, comprising:

- 10 a) obtaining embryogenic cells;
- b) exposing said cells to a heterologous DNA molecule under conditions whereby said DNA enters the plastids of said cells, said heterologous DNA molecule encoding at least one exogenous protein, said at least
- 15 one exogenous protein encoding a selectable marker;
- c) applying a selection agent to said cells to facilitate sorting of untransformed plastids from transformed plastids, said cells containing transformed plastids surviving and dividing in the presence of said
- 20 selection agent;
- d) transferring said surviving cells to selective media to promote shoot regeneration and growth; and
- e) rooting said shoots, thereby producing
- 25 transplastomic monocot plants.

19. A method as claimed in claim 18, wherein said heterologous DNA molecule is introduced into said plant cell via a process selected from the group consisting of

30 biolistic bombardment, Agrobacterium-mediated transformation, microinjection and electroporation.

20. A method as claimed in claim 18, wherein protoplasts are obtained from said embryogenic cells and said heterologous DNA molecule is delivered to said protoplasts by exposure to polyethylene glycol.

5

21. A method as claimed in claim 18, wherein said selection agent is selected from the group consisting of streptomycin, and paromomycin

10

22. A monocot transformed via the method of claim 18.

23. A transformed monocot plant as claimed in claim 22, said monocot plant being selected from the group consisting of maize, millet, sorghum, sugar cane, rice, wheat, barley, oat, rye, and turf grass.

15

24. A method for producing transplastomic rice plants, said method comprising:

20

a) obtaining embryogenic calli;

b) inducing proliferation of calli on modified CIM medium;

c) obtaining embryogenic cell suspensions of said proliferating calli in liquid AA medium;

25

d) bombarding said embryogenic cells with microprojectiles coated with plasmid DNA;

e) transferring said bombarded cells to selective liquid AA medium;

30

f) transferring said cells surviving in AA medium to selective RRM regeneration medium for a time period sufficient for green shoots to appear; and

g) rooting said shoots in a selective MS salt medium.

25. A method as claimed in claim 24, said plasmid DNA being selected from the group of plasmids consisting of pMSK35 and pMSK53, pMSK54 and pMSK49.

26. A transplastomic rice plant produced by the method of claim 24.

10

27. A method for containing transgenes in transformed plants, comprising:

a) determining the codon usage in said plant to be transformed and in microbes found in association with said plant; and

b) genetically engineering said transgene sequence via the introduction of rare codons to abrogate expression of said transgene in said plant associated microbe.

20

28. A method as claimed in claim 27, wherein said transgene is a bar gene and said rare codons are arginine encoding codons selected from the group consisting of AGA and AGG.

25

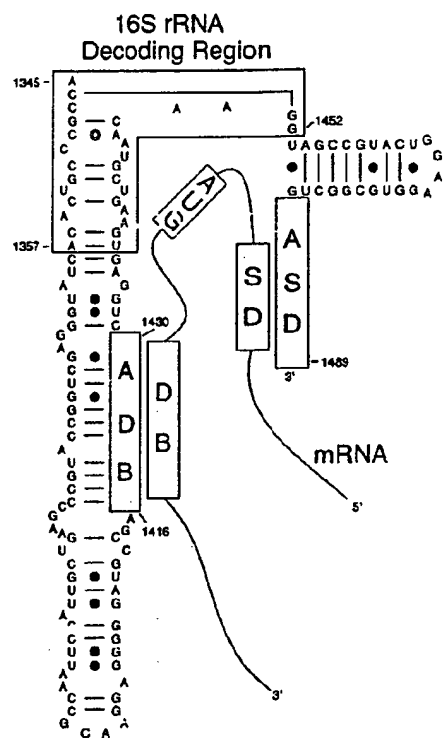


Figure 1A



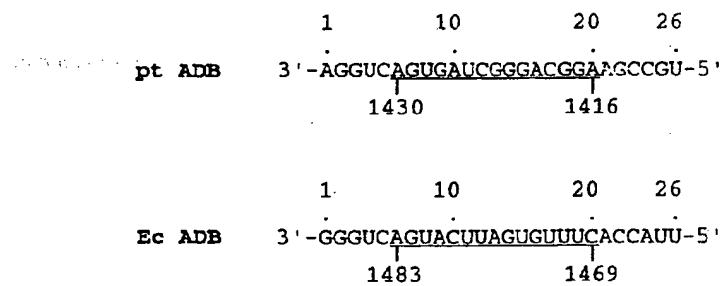


Figure 1B

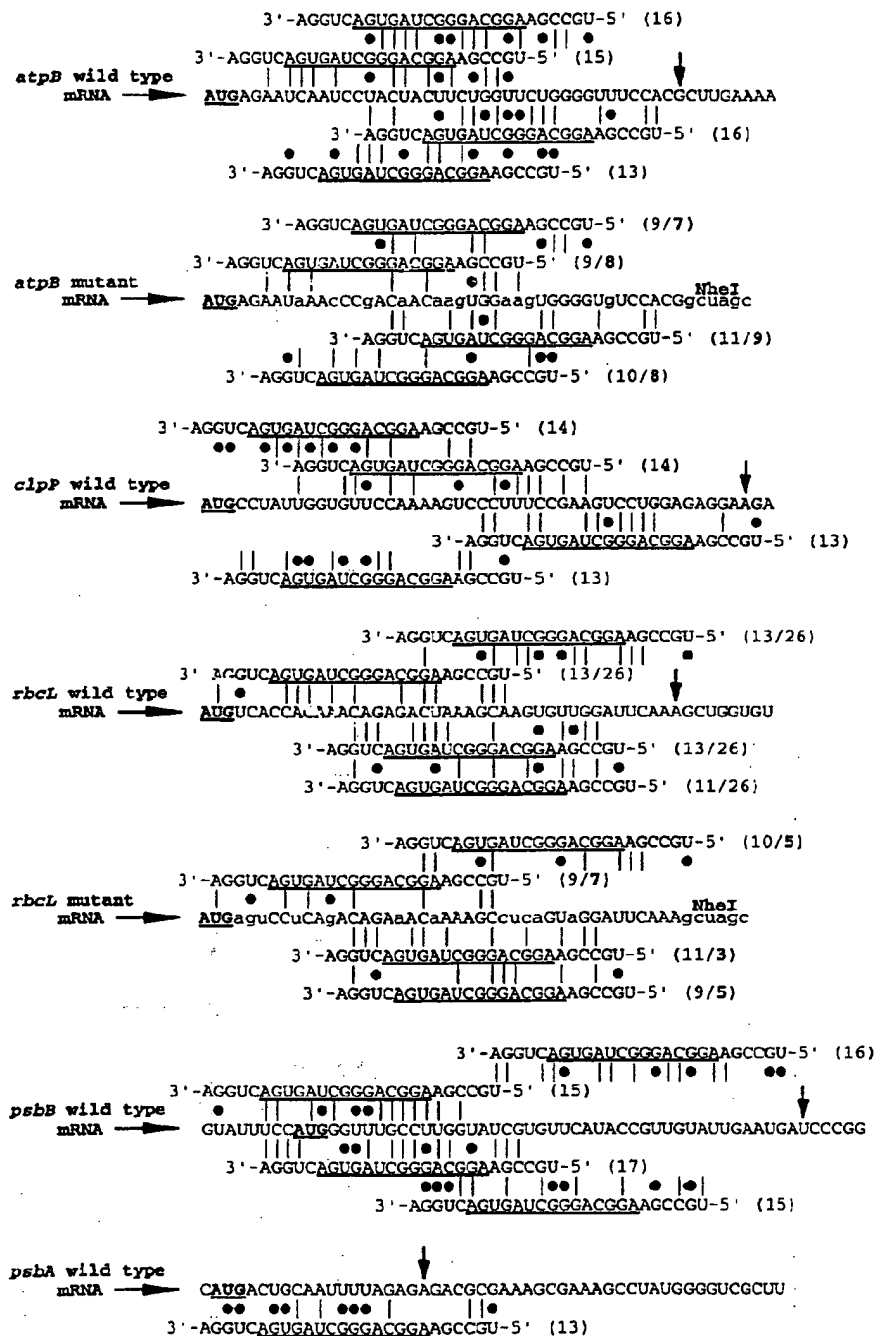


Figure 2A

T7g10 mRNA  
 AUGGCUAGCAUGACUGGUGGACAGCAAAUGGGUCGCGGAUCCGGCUGCUA  
 Ec ADB 3'-GGGUCAGUACUAGUGUUUACCAU-5' (15)

T7g10+DB/Ec mRNA  
 AUGGCAAGCAUGACUGGUGGACAGGCUagc  
 pt ADB 3'-AGGUCAGUGAUCCGGACCGGAAGCCGU-5' (13)

T7g10+DB/pt mRNA  
 AUGGCaaAucacuaagcccugccUUGGCGagc  
 pt ADB 3'-AGGUCAGUGAUCCGGACCGGAAGCCGU-5' (21)

T7g10-DB mRNA  
 ACAUAUUGGCUagcauguuagcaaggaugaucau  
 pt ADB 3'-AGGUCAGUGAUCCGGACCGGAAGCCGU-5' (14)

Figure 2B

**SacI**

**SacI**

**SacI**

**SacI**

**SacI**

Figure 3A

**PrrnLrbcL+DBwt (pHK14)**

SacI  
 1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
 51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG TCGAGTAGAC  
 101 CTTGTTGTTG TGAA AATTCT TAATTCATGA GTTGTAGGA GGGATTTATG  
 NheI  
 151 TCACCACAAA CAGAGACTAA AGCAAGTGTG GGATTCAAAG ctagc

**PrrnLrbcL-DB (pHK15)**

SacI  
 1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
 51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG TCGAGTAGAC  
 101 CTTGTTGTTG TGAA AATTCT TAATTCATGA GTTGTAGGA GGGATTTATG  
 NheI  
 151 TCAGctagc

**PrrnLrbcL+DEm (pHK54)**

SacI  
 1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
 51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG TCGAGTAGAC  
 101 CTTGTTGTTG TGAA AATTCT TAATTCATGA GTTGTAGGA GGGATTTATG  
 NheI  
 151 aguCCuCAgA CAGAA CaAA AGCcucaGTa GGATTCAAAG ctagc

**PrrnLpsbB+DBwt (pHK16)**

SacI  
 1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
 51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG CAATGCAATA  
 101 AAGTTACGTA GTGTCTATTT ATCTTTGATA TAAGGGGTAT TTCCATGGGT  
 NheI  
 151 TTGCCTTGGT ATCGTGTTCa TACCGTTGTA TTGAATGATg ctagc

**PrrnLpsbB-DB (pHK17)**

SacI  
 1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
 51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG CAATGCAATA  
 NcoI NheI  
 101 AAGTTACGTA GTGTCTATTT ATCTTTGATA TAAGGGGTAT TTccatggct  
 151 agc

**Figure 3B**

1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAA AAAAGCCTTC  
101 CATT TTTCTAT TTTGATTGT AGAAACTAG TGTGCTTGGG AGTCCCTGAT  
151 GATTAAATA ACCAAGATT TACCATGACT GCAATTTTAG AGAGAgctag  
201 c

SacI  
1 gagctcGCTC CCCGCCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGA<sup>●</sup> AAAAGCCTTC  
101 CATTTTCTAT TTGATTGT AGAAACTAG TGTGCTTGGG AGTCCCTGAT  
NcoI NheI  
151 GATTAAATAA ACCAAGATT TAccatggct agc

SacI  
1 gagctcGCTC CCCCGCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGC  
51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG CAAAAAGCCT  
101 TCCATTTTCT ATTTTGATT GTAGAAACT AGTGTGCTTG GGAGTCCCTG  
NcoI NheI  
151 ATGATTAAAT AAACCAAGAT TTTAccatgg ctatgc

**Figure 3C**

**PrrnLT7g10+DB/Ec (pHK18)**

SacI  
1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG GGAGACCACA  
101 ACGGTTTCCC aCTAGAAATA ATTTTGTTTA ACTTTAAGAA GGAGATATAC  
NheI  
151 ATATGGCaAG CATGACTGGT GGACAGgcta gc

**PrrnLT7g10+DB/pt (pHK19)**

SacI  
1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG GGAGACCACA  
101 ACGGTTTCCC aCTAGAAATA ATTTTGTTTA ACTTTAAGAA GGAGATATAC  
NheI  
151 ATATGGCaAt cactagccct gccttGgcta gc

**PrrnLT7g10-DB (pHK20)**

SacI  
1 gagctcGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG GGAGACCACA  
101 ACGGTTTCCC aCTAGAAATA ATTTTGTTTA ACTTTAAGAA GGAGATATAC  
NheI  
151 ATATGgctag c

**Figure 3D**

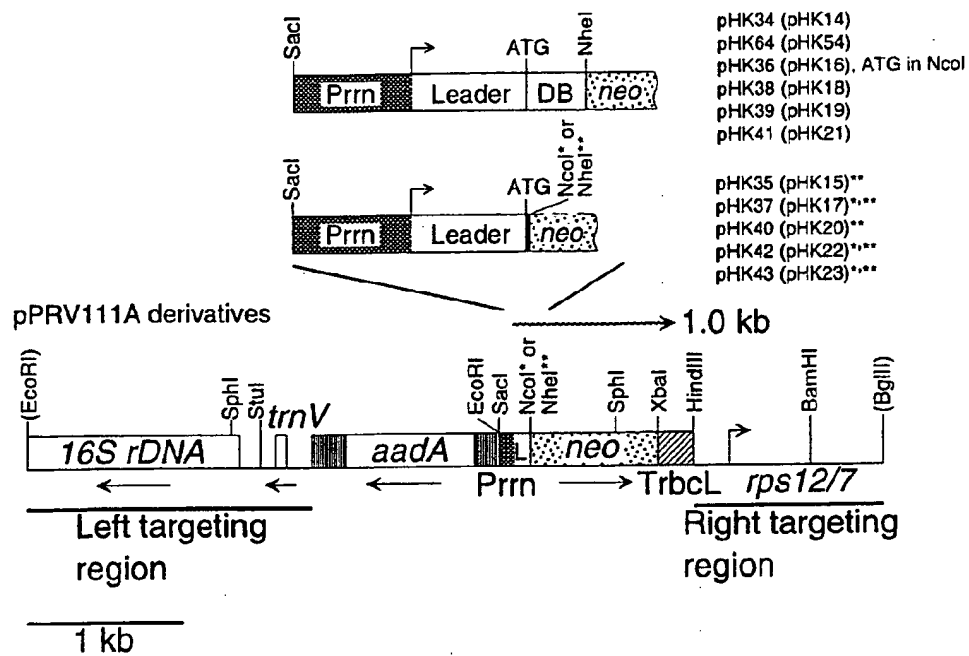


Figure 4A



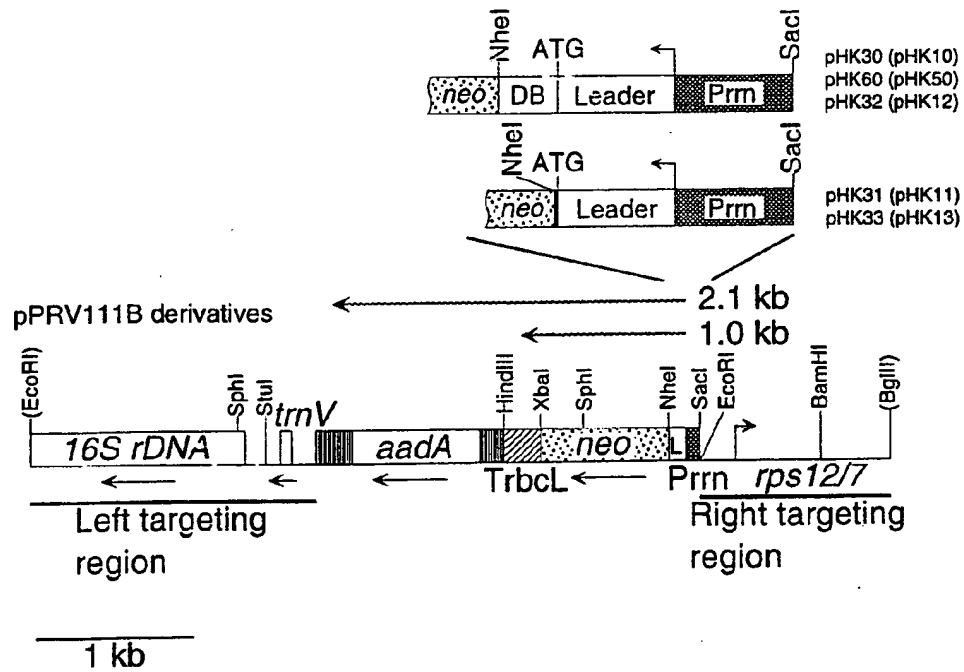


Figure 4B

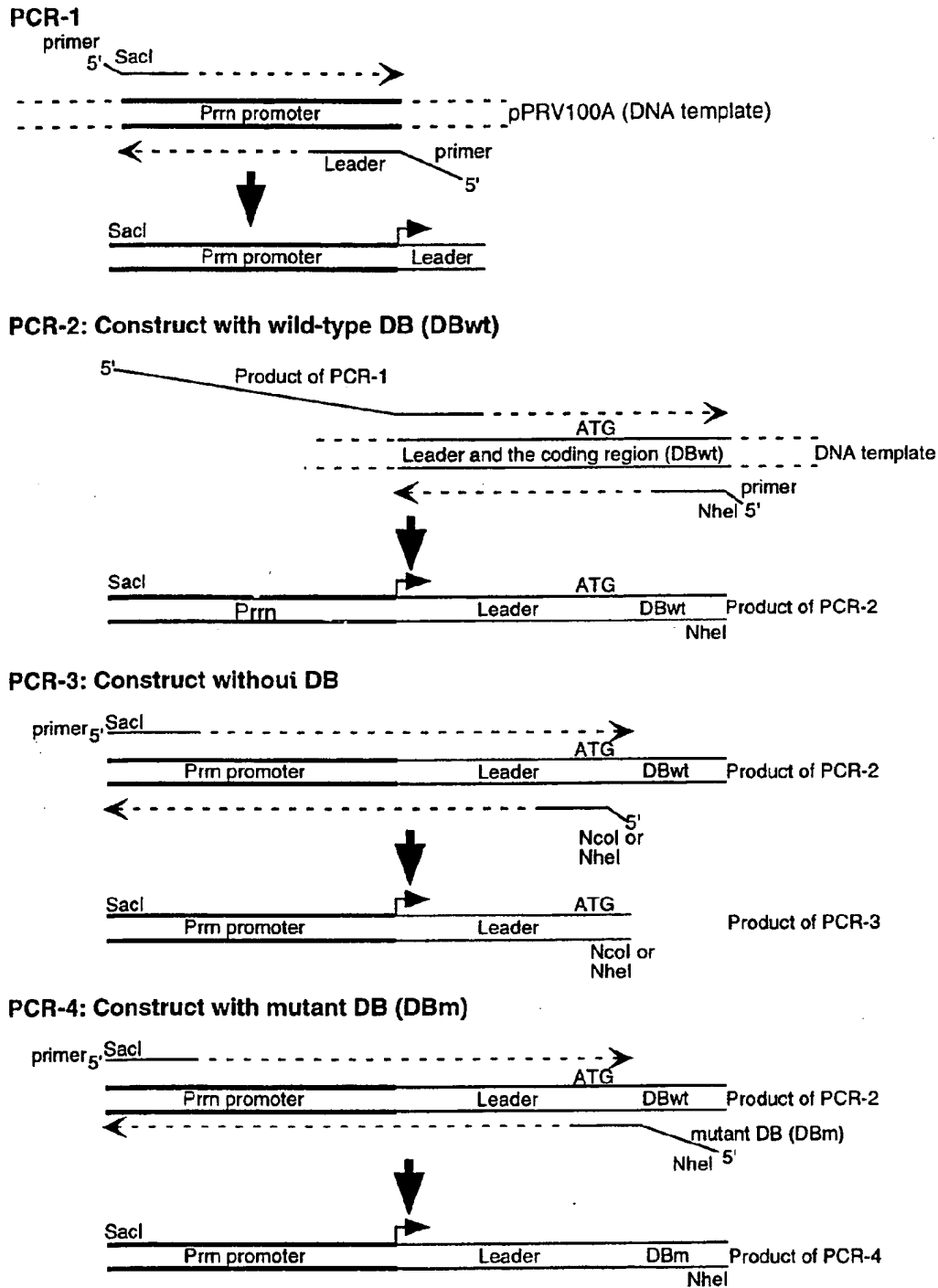


Figure 5

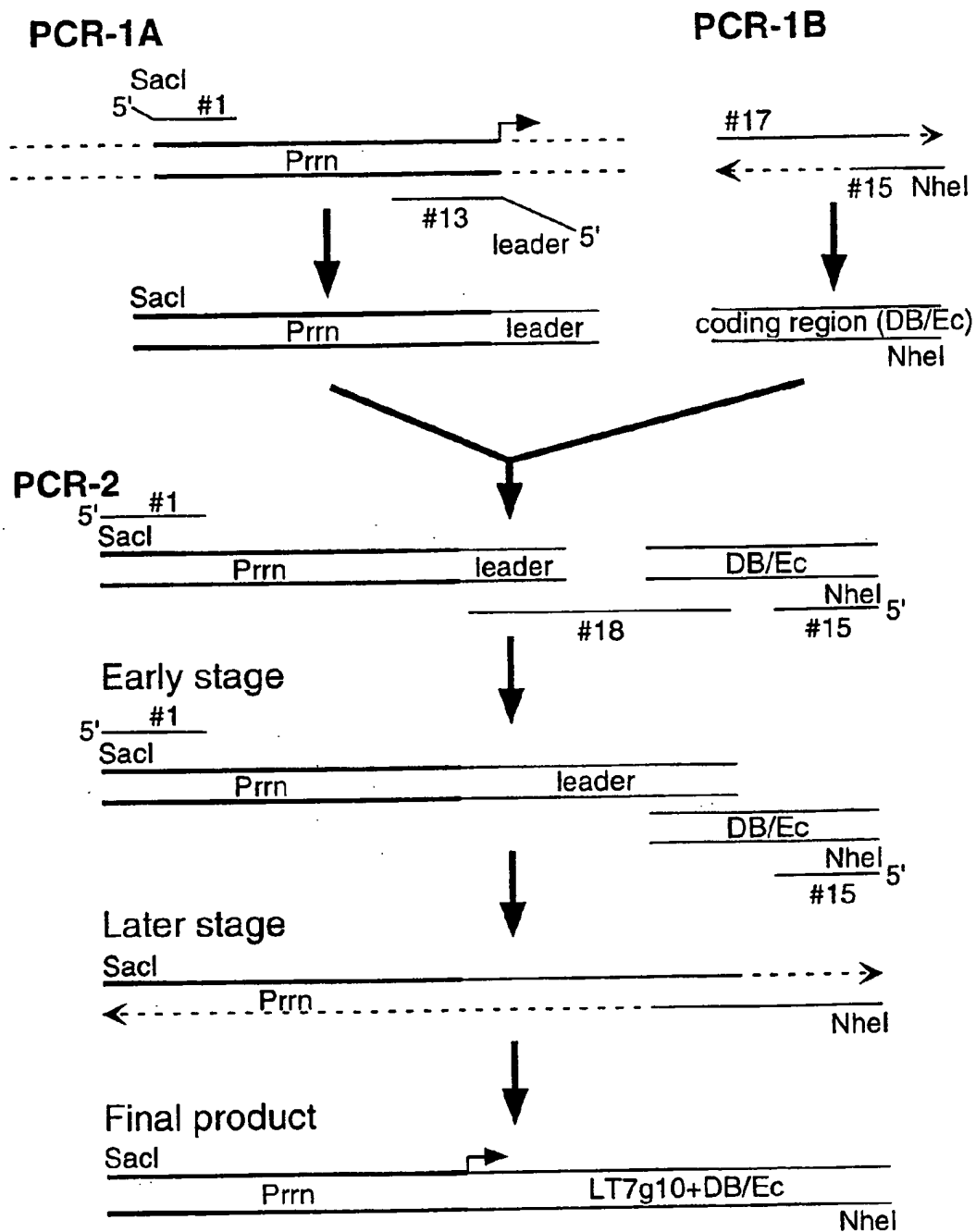


Figure 6

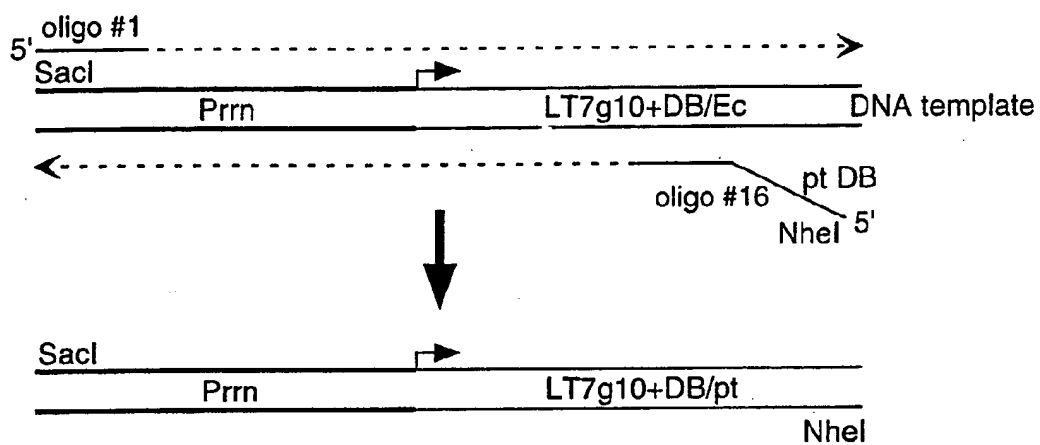


Figure 7

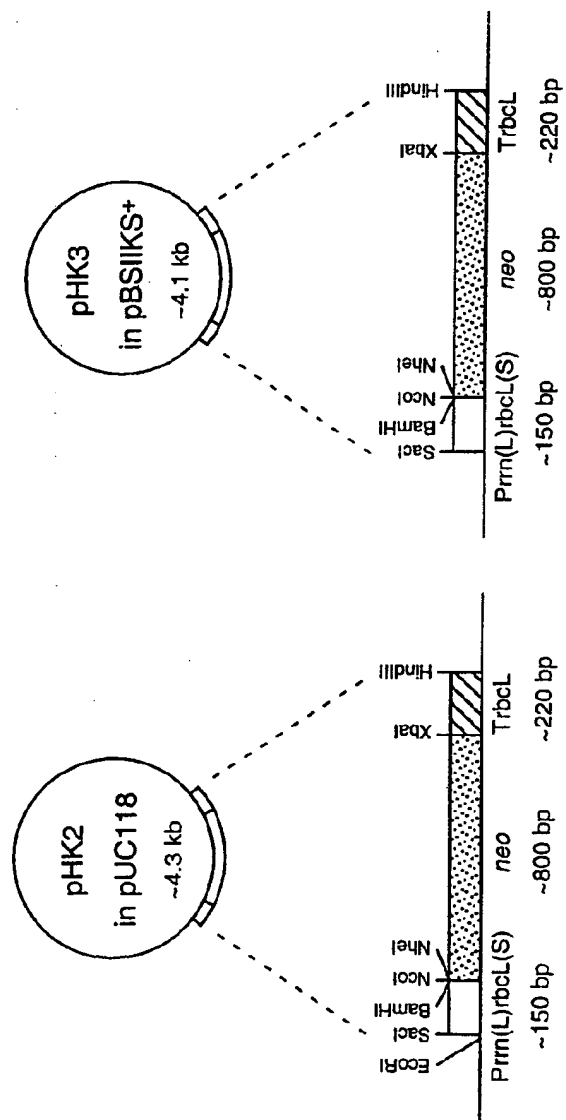


Figure 8

```

      SacI
1   gagctcggta cccaaaGCTC CCCC GCCGTC GTTCAATGAG AATGGATAAG
51  AGGCTCGTGG GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG
101 CGAACTCCGG GCGAATAcGA AGCGctTGA TACAGTTGTA GGGAGGGATc
      NheI
151 catggctagc ATTGAACAAG ATGGATTGCA CGCAGGTTCT CCGGCCGCTT
201 GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC
251 TCTGATGCCG CCGTGTTCGG GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT
301 TGTCAAGACC GACCTGTCCG GTGCCCTGAA TGAAGTCCAG GACGAGGCAG
351 CGCGGCTATC GTGGCTGGCC ACGACGGGCG TTCCTTGCGC AGCTGTGCTC
401 GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG GCGAAGTGCC
451 GGGGCAGGAT CTCCTGTCA TCTACCTTGC TCCTGCCGAG AAAGTATCCA
501 TCATGGCTGA TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC
551 CCATTCGACC ACCAAGCGAA ACATCGCATC GAGCGAGCAC GTACTCGGAT
601 GGAAGCCGGT CTTGTTCGATC AGGATGATCT GGACGAAGAG CATCAGGGGC
651 TCGCGCCAGC CGAACTGTTC GCCAGGCTCA AGGCGCGCAT GCCCGACGGC
701 GAGGATCTCG TCGTGACACA TGGCGATGCC TGCTTGCCGA ATATCATGGT
751 GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG
801 CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT TGCTGAAGAG
851 CTTGGCGGCG AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC
901 TCCCGATTCTG CAGCGCATCG CCTTCTATCG CCTTCTTGAC GAGTTCTTCT
      XbaI
951 GAgcgggtct agagtAGACA TTAGCAGATA AATTAGCAGG AAATAAGAA
1001 GGATAAGGAG AAAGAACTCA AGTAATTATC CTTCGTTCTC TTAATTGAAT
1051 TGCAATTAAA CTCGGCCCCA TCTTTTACTA AAAGGATTGA GCCGAATACA
1101 ACAAAGATTC TATTGCATAT ATTTTGACTA AGTATATACT TACCTAGATA
      HindIII
1151 TACAAGATTT GAAATACAAA ATCTAGcaag ctt

```

Figure 9

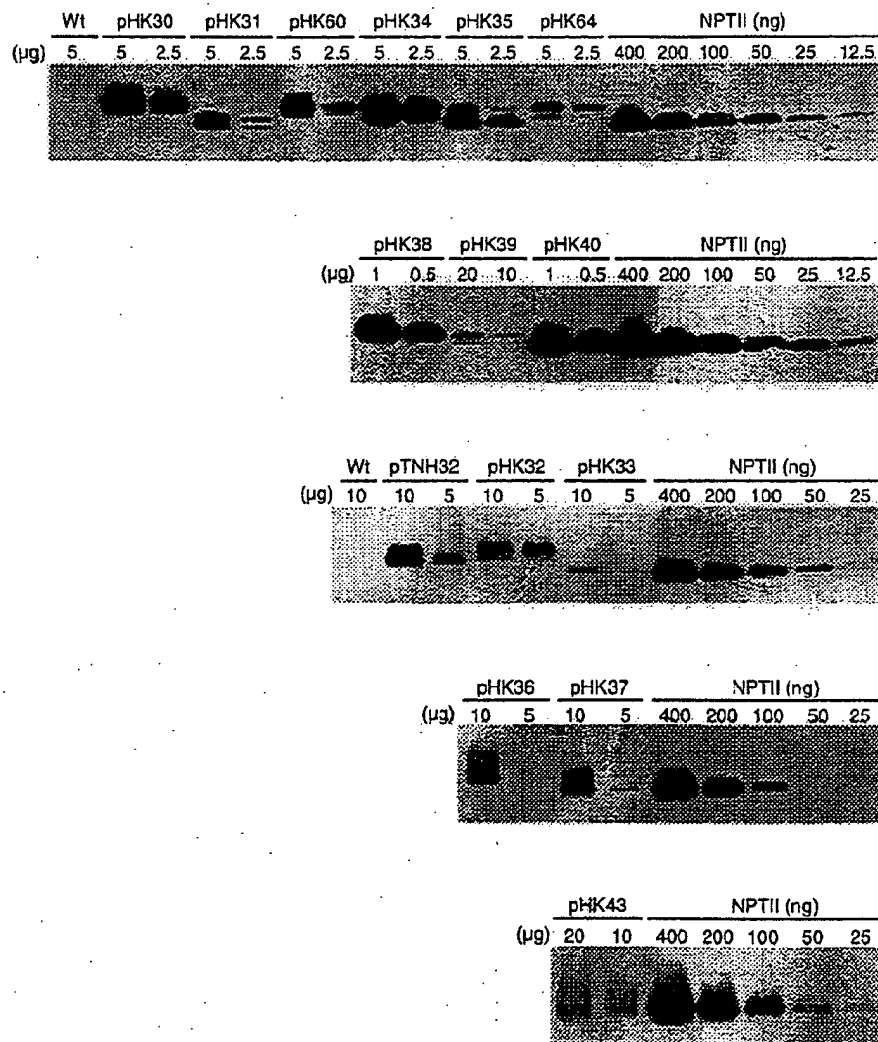


Figure 10

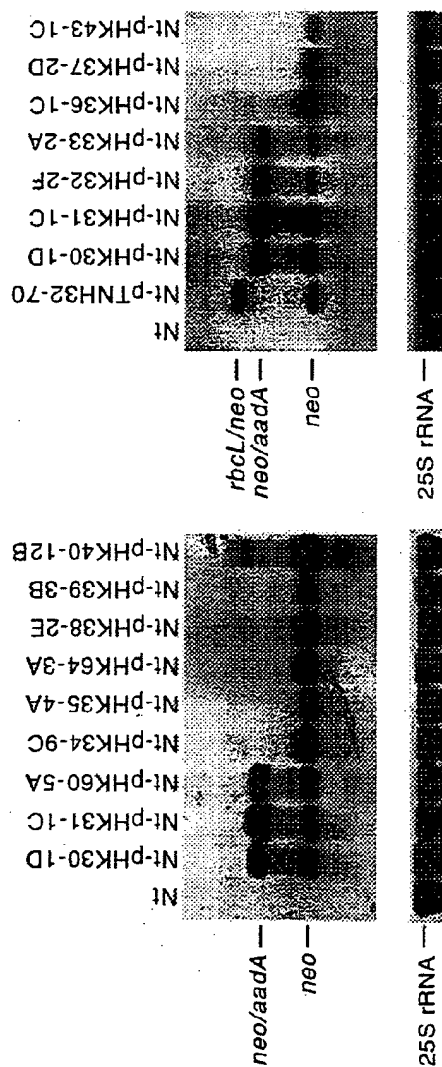


Figure 11



<i>atpB</i> wt	<b>AUG</b>	AGA	AUC	AAU	CCU	ACU	ACU	ACU	UCU	GGU	UCU	GGG	GUU	UCC	ACG
	Met	Arg	Ile	Asn	Pro	Thr	Thr	Thr	Ser	Gly	Ser	Gly	Val	Ser	Thr
	Fraction	1.0	0.22	0.27	0.61	0.30	0.37	0.37	0.31	0.38	0.31	0.26	0.35	0.14	0.15
Tripletlet/1000	24.6	7.8	15.5	18.1	13.5	18.4	18.4	20.2	28.2	20.2	19.2	24.9	9.1	7.5	
<i>atpB</i> m	<b>AUG</b>	AGA	AUA	AAC	CCG	ACA	ACA	agU	Gga	agU	GGG	GUG	UCC	ACG	
	Met	Arg	Ile	Asn	Pro	Thr	Thr	Ser	Gly	Ser	Gly	Val	Ser	Thr	
	Fraction	1.0	0.22	0.29	0.39	0.30	0.23	0.14	0.24	0.14	0.26	0.21	0.14	0.15	
Tripletlet/1000	24.6	7.8	16.6	11.4	13.2	11.7	11.7	9.3	17.9	9.3	19.2	15.3	9.1	7.5	
<i>rbcL</i> wt	<b>AUG</b>	UCA	CCA	CAA	ACA	GAG	ACU	AAA	GCA	AGU	GUU	GGA	UUC	AAA	
	Met	Ser	Pro	Gln	Thr	Glu	Thr	Lys	Ala	Ser	Val	Gly	Phe	Lys	
	Fraction	1.0	0.21	0.24	0.57	0.23	0.38	0.37	0.60	0.29	0.14	0.35	0.24	0.40	0.60
Tripletlet/1000	24.6	13.5	10.6	21.0	11.7	12.4	18.4	22.0	18.1	9.3	24.9	17.9	22.5	22.0	
<i>rbcL</i> m	<b>AUG</b>	agu	CCu	CAG	ACA	GAA	ACA	AAA	GCC	uca	GUA	GGA	UUC	AAA	
	Met	Ser	Pro	Gln	Thr	Glu	Thr	Lys	Ala	Ser	Val	Gly	Phe	Lys	
	Fraction	1.0	0.14	0.30	0.43	0.23	0.62	0.23	0.60	0.16	0.21	0.31	0.24	0.40	0.60
Tripletlet/1000	24.6	9.3	13.5	15.5	11.7	20.7	11.7	22.0	10.1	13.5	21.8	17.9	22.5	22.0	
T7g10+DB/EC	<b>AUG</b>	Gca	AGC	AUG	ACU	GGU	GGU	GGA	CAG	gcu	agc	auu	gaa	caa	gau
	Met	Ala	Ser	Met	Thr	Gly	Gly	Gln	Ala	Ser	Ile	Glu	Gln	Asp	
	Fraction	1.0	0.29	0.07	1.00	0.37	0.38	0.24	0.43	0.39	0.07	0.45	0.62	0.57	0.75
Tripletlet/1000	24.6	18.1	4.7	24.6	18.4	28.2	17.9	15.5	24.4	4.7	25.9	20.7	21.0	24.6	
T7g10+DB/pt	<b>AUG</b>	GCa	Auc	acu	agc	ccu	gcc	uug	gcu	agc	auu	gaa	caa	gau	
	Met	Ala	Ile	Thr	Ser	Pro	Ala	Leu	Ala	Ser	Ile	Glu	Gln	Asp	
	Fraction	1.0	0.29	0.27	0.37	0.07	0.30	0.16	0.24	0.39	0.07	0.45	0.62	0.57	0.75
Tripletlet/1000	24.6	18.1	15.5	18.4	4.7	13.5	10.1	34.7	24.4	4.7	25.9	20.7	21.0	24.6	
T7g10-DB	<b>AUG</b>	gcu	agc	auu	gaa	caa	gau	gga	uug	cac	gca	ggu	ucu	ccg	
	Met	Ala	Ser	Ile	Glu	Gln	Asp	Gly	Leu	His	Ala	Gly	Ser	Pro	
	Fraction	1.0	0.39	0.07	0.45	0.62	0.57	0.75	0.24	0.28	0.29	0.38	0.31	0.30	
Tripletlet/1000	24.6	24.4	4.7	25.9	20.7	21.0	24.6	17.9	34.7	9.1	18.1	28.2	20.2	13.2	

Figure 12

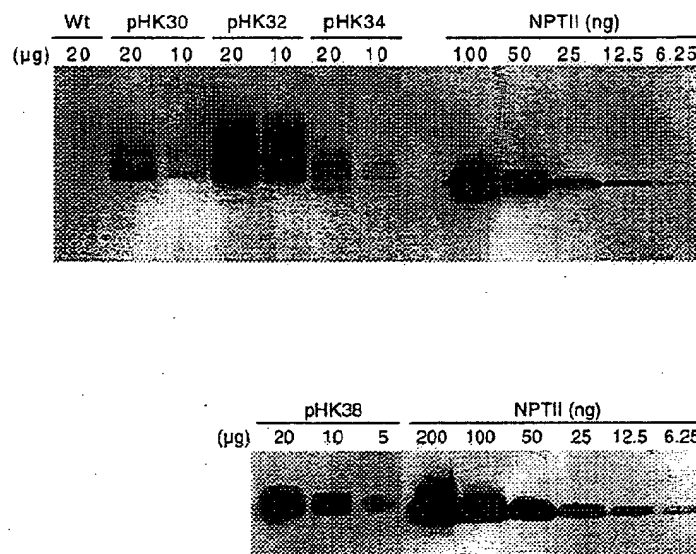


Figure 13A

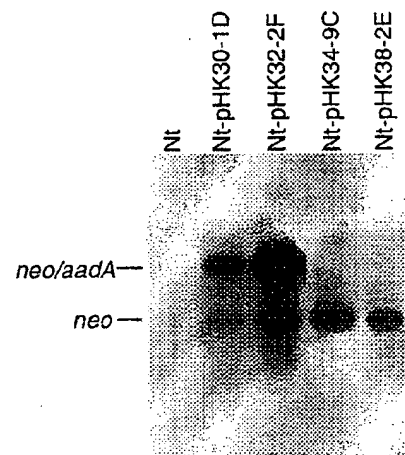


Figure 13B

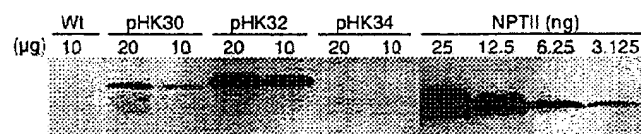


Figure 14

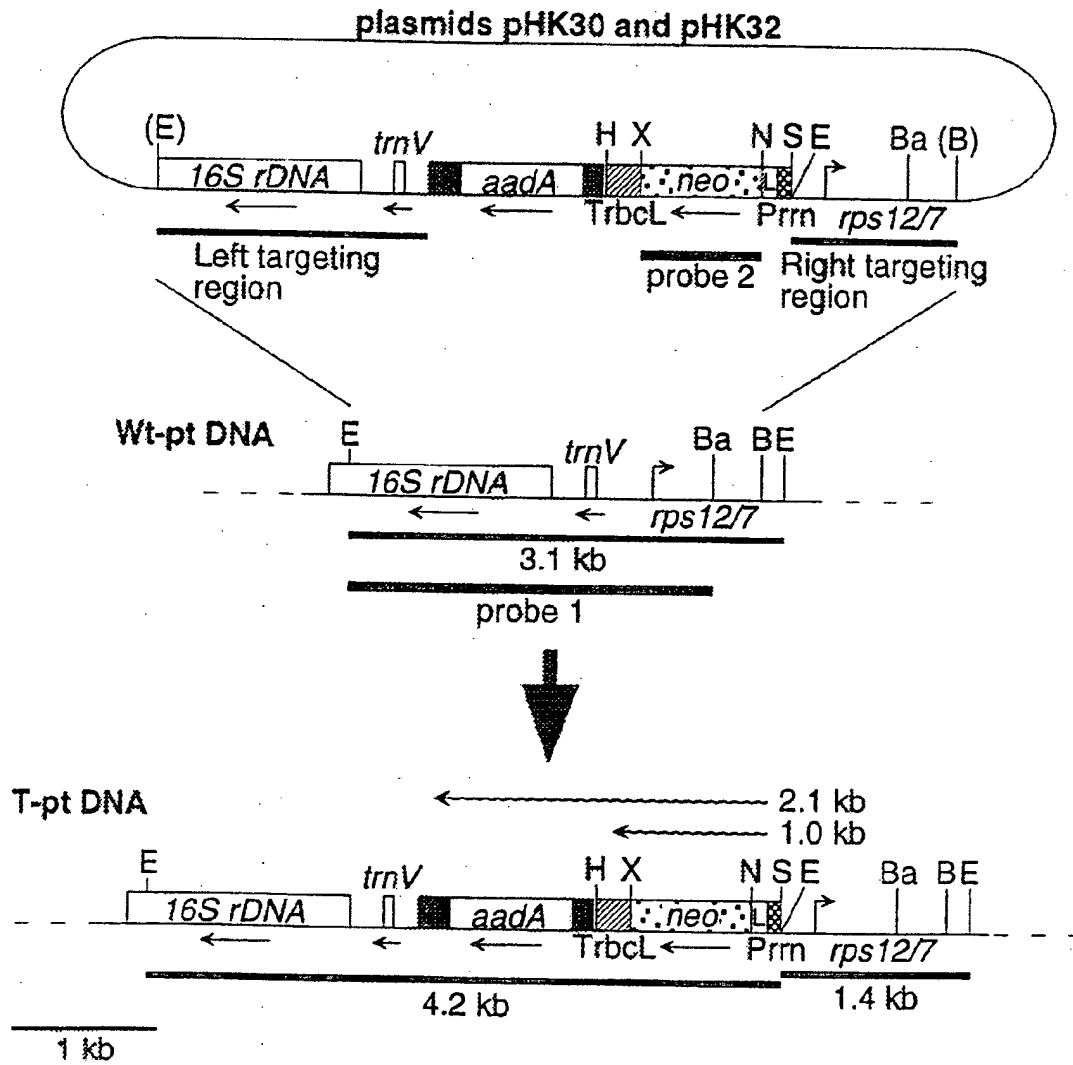


Figure 15A

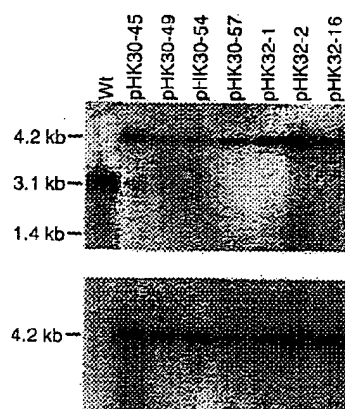


Figure 15B

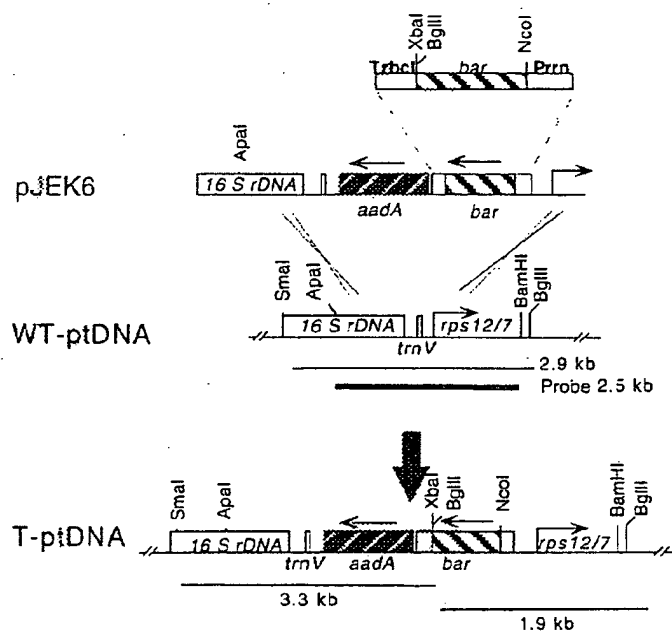


Figure 16A

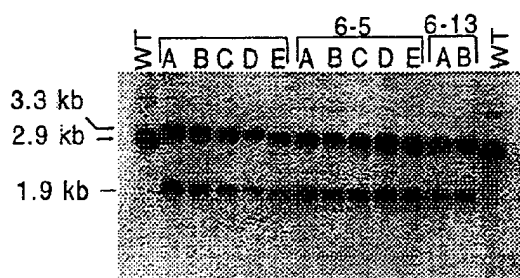


Figure 16B



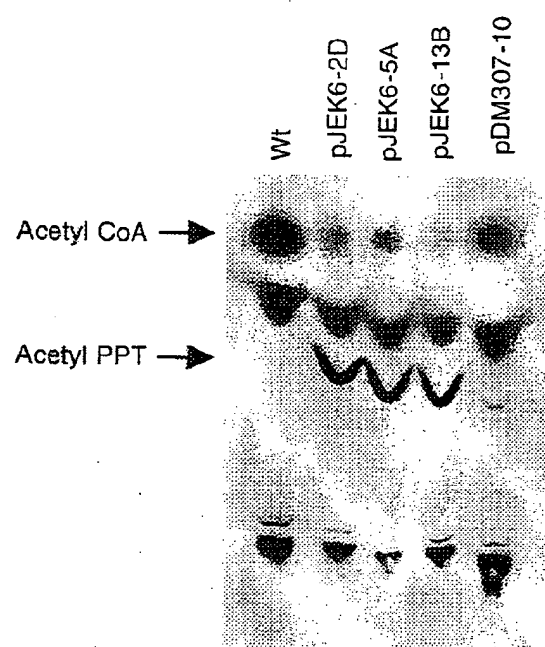
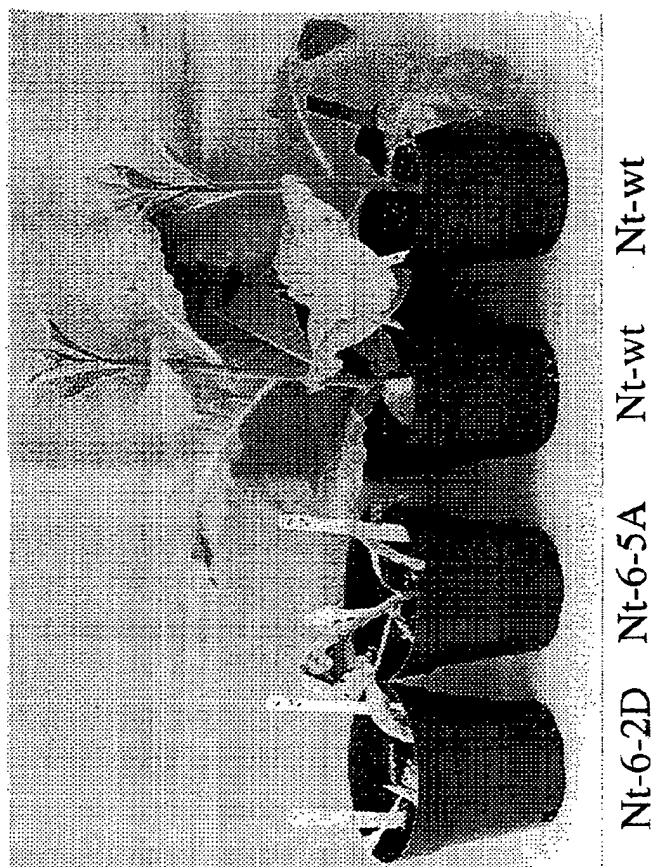


Figure 17

**Figure 18A**

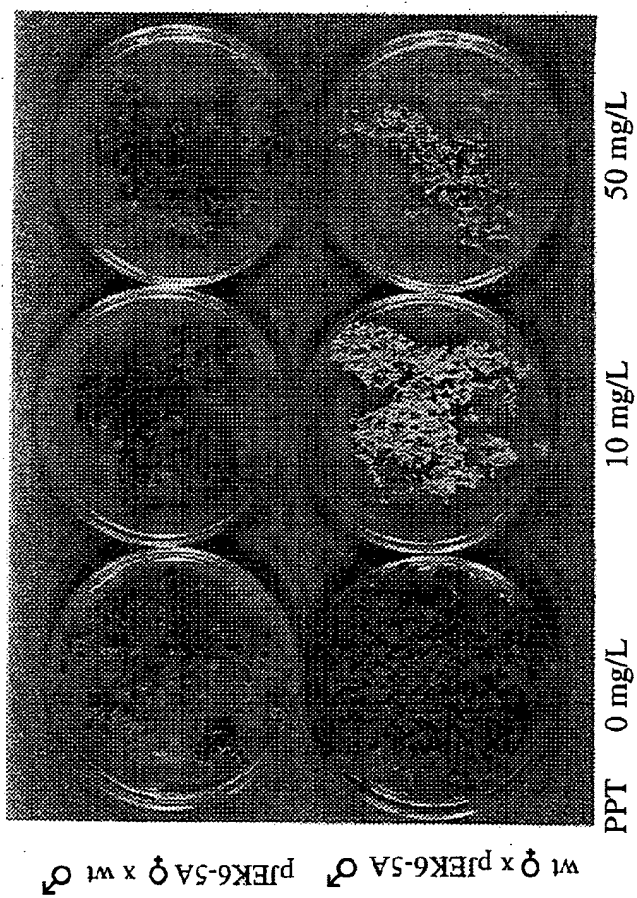


Figure 18B

**NcoI**

CCATGgcaccacaaacagagAGCCCAGAACGACGCCCGCCGACATCCGCCGTGCCACCG  
-----+-----+-----+-----+-----+ 60  
CGTACcgtggtgtttgtctcTCGGGTCTTGCTGCGGGCCGGCTGTAGGCGGCACGGTGGC  
M A P Q T E S P E R R P A D I R R A T E  
AGGCGGACATGCCGGCGGTCTGCACCATCGTCAACCACTACATCGAGACAAGCACGGTCA  
-----+-----+-----+-----+-----+ 120  
TCCGCCTGTACGGCCGCCAGACGTGGTAGCAGTTGGTGATGTAGCTCTGTTTCGTGCCAGT  
A D M P A V C T I V N H Y I E T S T V N  
ACTTCCGTACCGAGCCGAGGAACCGCAGGAGTGGACGGACGACCTCGTCCGTCTGCGGG  
-----+-----+-----+-----+-----+ 180  
TGAAGGCATGGCTCGGCGTCTTGCGCTCACCTGCCTGCTGGAGCAGGCAGACGCCC  
F R T E P Q E P Q E W T D D L V R L R E  
AGCGCTATCCCTGGCTCGTCGCCGAGGTGGACGGCGAGGTGCGCCGGCATCGCCTACGCGG  
-----+-----+-----+-----+-----+ 240  
TCGCGATAGGGACCGAGCAGCGGCTCCACCTGCCGCTCCAGCGGCCGTAGCGGATGCGCC  
R Y P W L V A E V D G E V A G I A Y A G  
GCCCCTGAAGGCACGCAACGCCTACGACTGGACGGCCGAGTCGACCGTGTACGTCTCCC  
-----+-----+-----+-----+-----+ 300  
CGGGGACCTTCCGTGCGTTGCGGATGCTGACCTGCCGGCTCAGCTGGCACATCAGAGGG  
P W K A R N A Y D W T A E S T V Y V S P  
CCCCGCCACCAGCGGACGGGACTGGGCTCCACGCTCTACACCCACCTGCTGAAGTCCCTGG  
-----+-----+-----+-----+-----+ 360  
GGGCGGTGGTTCGCTGCCCTGACCCGAGGTGCGAGATGTGGGTGGACGACTTCAGGGACC  
R H Q R T G L G S T L Y T H L L K S L E  
AGGCACAGGGCTTCAAGAGCGTGGTTCGCTGTATCGGGCTGCCCAACGACCCGAGCGTGC  
-----+-----+-----+-----+-----+ 420  
TCCGTGTCCCGAAGTTCTCGCACCAGCGACAGTAGCCCGACGGGTTGCTGGGCTCGCAGC  
A Q G F K S V V A V I G L P N D P S V R  
GCATGCACGAGGCGCTCGGATATGCCCCCGCGGCATGCTGCGGGCGGCGGCTTCAAGC  
-----+-----+-----+-----+-----+ 480  
CGTACGTGCTCCGCGAGCCTATACGGGGGGCGCCGTACGACGCCCCGCGGCCGAAGTTCG  
M H E A L G Y A P R G M L R A A G F K H  
ACGGGAACCTGGCATGACGTGGGTTTCTGGCAGCTGGACTTCAGCCTGCCGGTACCGCCCC  
-----+-----+-----+-----+-----+ 540  
TGCCCTTGACCGTACTGCACCCAAAGACCGTCGACCTGAAGTCGGACGGCCATGGCGGGG  
G N W H D V G F W Q L D F S L P V P P R

**BglII**

GTCCGGTCTGCCCCGTACCGAGATCTGATGatcgaattcctgcagcccggggatccac  
-----+-----+-----+-----+-----+ 600  
CAGGCCAGGACGGGCAGTGGCTCTAGACTACTagcttaaggacgtcgggccccctaggtg  
P V L P V T E I \*

**XbaI**

tagttctaga  
-----+ 610  
atcaagatct

**Figure 19**

**NcoI NheI**

CcATGgctAGCCCAGAAAGaGACCGGCCGAATtaGaCGTGCTACaGAAGctGAtATGC  
 -----+-----+-----+-----+-----+  
 ggTACcgaTCGGGTCTTtCTtCtGGCCGGCTaTAatCtGCACGaTGtCTtCGaCTaTACG  
 M A S P E R R P A D I R R A T E A D M P

CaGCaGTTtGTACaATtGTtAATCAtTATaTAaACAAGtACcGTaAACTTtCgaACTg  
 -----+-----+-----+-----+-----+  
 GtCGtCAaACaTGTtAaCAaTTaGTaATaTATcTtTGTTCaTGgCATTTGAAagCtTGaC  
 A V C T I V N H Y I E T S T V N F R T E

AaCctCAaGAACcCAaGAaTGGACTGAAGattTaGTCCGTtTaCGaGAGCGCTATCCtT  
 -----+-----+-----+-----+-----+  
 TtGGaGTTcTTGGaGTTcTtACCTGaCTaCTaaAtCAGGCAaAtGctCTCGCGATAGGAa  
 P Q E P Q E W T D D L V R L R E R Y P W

GGCTtGTaGCaGAaGTtGACGGaGAaGTaGCTGGgATtGCaTATGCGGGCCCGTGGAAaG  
 -----+-----+-----+-----+-----+  
 CCGAaCATCGtCTtCAaCTGCCtCTtCATCGaCCcTaaCGtATaCGCCCGGGcACCTTtC  
 L V A E V D G E V A G I A Y A G P W K A

CacGaAATGCaTATGAtTGGACgGCTGAaTCAACTGTgTACGTtTCaCCaCGtCATCAaC  
 -----+-----+-----+-----+-----+  
 GTgCtTTaCGtATaCTaACCTGcCGaCTtAGtTGaCacATGCAaAGtGGtGCaGTaGTtG  
 R N A Y D W T A E S T V Y V S P R H Q R

GgCaGGAActTGGtTCTaCtTtTaTATaCCcATCTaCTGAaATCtTtTGGAGGCACAgGGtT  
 -----+-----+-----+-----+-----+  
 CcTGTcCTGAaCCaAGaTGaaAtATaTGgGTaGAtGACTTtAGaaACCTCCGTGTtCCCaA  
 T G L G S T L Y T H L L K S L E A Q G F

TtAAGAGtGTgGTaGCTGTtATaGGatTGCCgAATGAtCCctcgGTaCGCATGCAcGAaG  
 -----+-----+-----+-----+-----+  
 AaTTCTCaCACaTcGACAAaTATCCtaACGGcTTaCTaGGgagcCATGCGTACGTgCTtC  
 K S V V A V I G L P N D P S V R M H E A

CtCTcCGATATGCTCCCaGaGGtATGtTGaGGGCcGCAGgtTTCAAaCATGGaAATGGC  
 -----+-----+-----+-----+-----+  
 GaGAgCCTATACGaGGgtCtCCaTACaACTCCCGgCGtCCaAAGTTtGTaCCTTTaACCG  
 L G Y A P R G M L R A A G F K H G N W H

ATGAtGTaGGTTTtTGGCAaCTtGAcTTCTcttTaCCaGTACcTCCtCGTCCcGTtTaC  
 -----+-----+-----+-----+-----+  
 TACTaCATCCAAaAACCGTtGAaCTgAAGagaaAtGGtCATGGaGGaGCAGGgCAaaAtG  
 D V G F W Q L D F S L P V P P R P V L P

**BglIII****XbaI**

CcGTTACTGAGATCTGATGAtctaga  
 -----+-----+-----+-----+-----+

GgCAaTGaCTCTAGACTACTagatct

V T E I \* \*

Figure 20A

NcoI NheI

ccATGgctAGCCGAGAAaGAaGACCGGCCGAtATtaGaCGTGctACaGAaGctGAtATGC  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ggTACcgaTCGGGTCTTtCTtCtGGCCGGCTaTAatCtGCACGaTgtCtTCGaCTaTACG  
 M A S P E R R P A D I R R A T E A D M P

CaGCaGtTtGTaCaAtTtGTtAATCATtTAtATaGAaACAAGtACaGTaAAAtTTtcGaACtG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GtCGtCAaACaTgtTAaCAaTTaGTaATaTAtCTtTGTTCaTgtCATtTAaAagCtiGac  
 A V C T I V N H Y I E T S T V N F R T E

AaCCtCAaGAACctCAaGAaTGGACtGAtGAttTaGTaCGTtTaCGaGAaCGtTATCctT  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TtGGAGtTCTTGGAGtTCTtACCTGaCTaCTaaAtCATGCAaAtGCTcTtGCaATAGGaA  
 P Q E P Q E W T D D L V R L R E R Y P W

GGCTtGTaGCaGAaGtTtGAcGGAaGtAGctGGaATtGCaTATGctGGtCCgTGGAAaG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CCGAaCAtCGtCTtCAaCTgCCTtCTtCATCGaCCTTAaCGtATaCGaCCaGGcACCTTtC  
 L V A E V D G E V A G I A Y A G P W K A

CACGaAAtGCaTATGAtTGGACaGCTGAaTCAActGtTtTATGtTtCaCCaCGtCATCAaC  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GTgCtTtFaCGtATaCTaACCTGtCGaCTtAGtTGaCAaATaCAaAGtGGtGCAGtAGtG  
 R N A Y D W T A E S T V Y V S P R H Q R

GtACaGGAActtGGtTCTtACtTtTATaTACtCATCTtCTtAAaTCtTGGaAGCACaAGtT  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CaTGTCTTGAAcCaAGaTGaaAtATaTGaGTaGAaGAaTtTAGaaACCTtCGTGTtCCaA  
 T G L G S T L Y T H L L K S L E A Q G F

TtAAaAGtGTaGTaGCTGtTATaGGatTGCCgAAAtGAtCCctcAGTaCGCATGCATGAaG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AaTTtTCaCAtCATCGACaATATCCtaACGGcTTaCTaGGgagtCATGCGTACGTaCTtC  
 K S V V A V I G L P N D P S V R M H E A

CtCTtGGATATGCTCCCaGaGGtATGtTGaGGGCaGCaGGtTTCRAaCAtGGaAAtTGGC  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GaGAaCCTATACGaGGgtCtCCaTACaActCCCGtCGtCCaAGTtTtGtACctTtTAACCG  
 L G Y A P R G M L R A A G F K H G N W H

ATGAtGTaGGTTTTtGGCAaCTtGAcTTCTtTtTaCCaGTACCTCCtCGTCCCGtTtTaC  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TACTaCAtCCAAaAACCGtTGAaCTgAAGagaaAtGGtCATGGaGGaGCAGGgCAaaAtG  
 D V G F W Q L D F S L P V P P R P V L P

BglIII XbaI

CcGTtACTGAGATCTGATGAtctaga

GgCAaTGaCTCTAGACTACTagatct  
 V T E I \* \*

Figure 20B

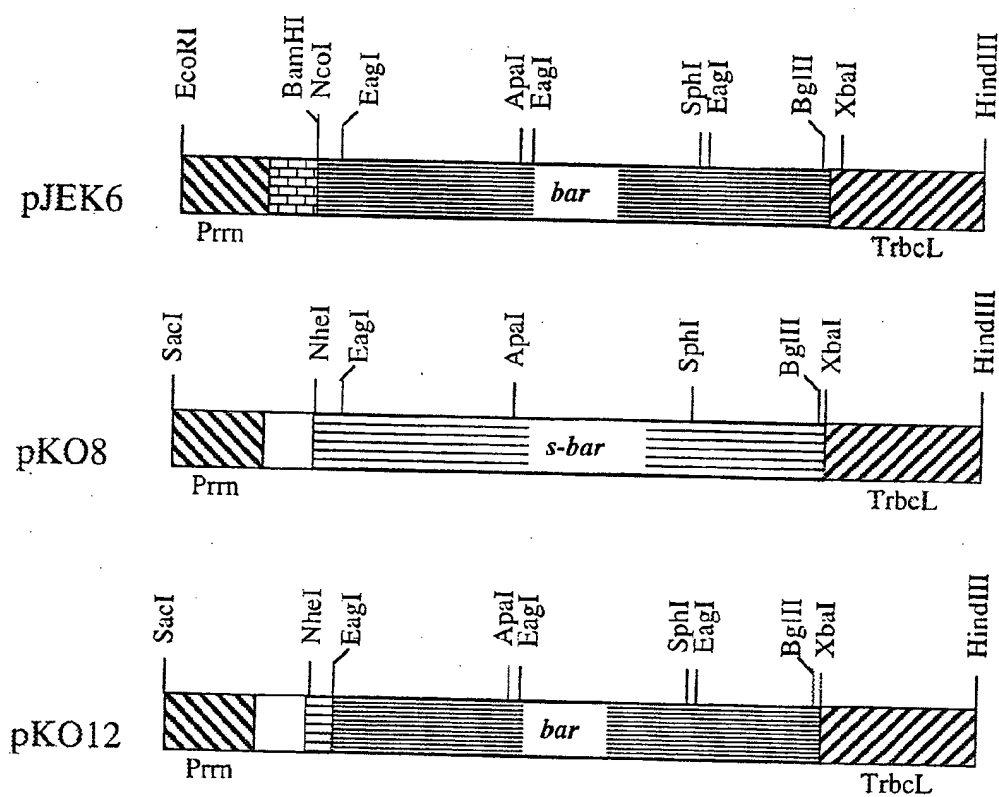


Figure 21

## Bacterial Extracts



Figure 22A



## Plant Extracts



Figure 22B

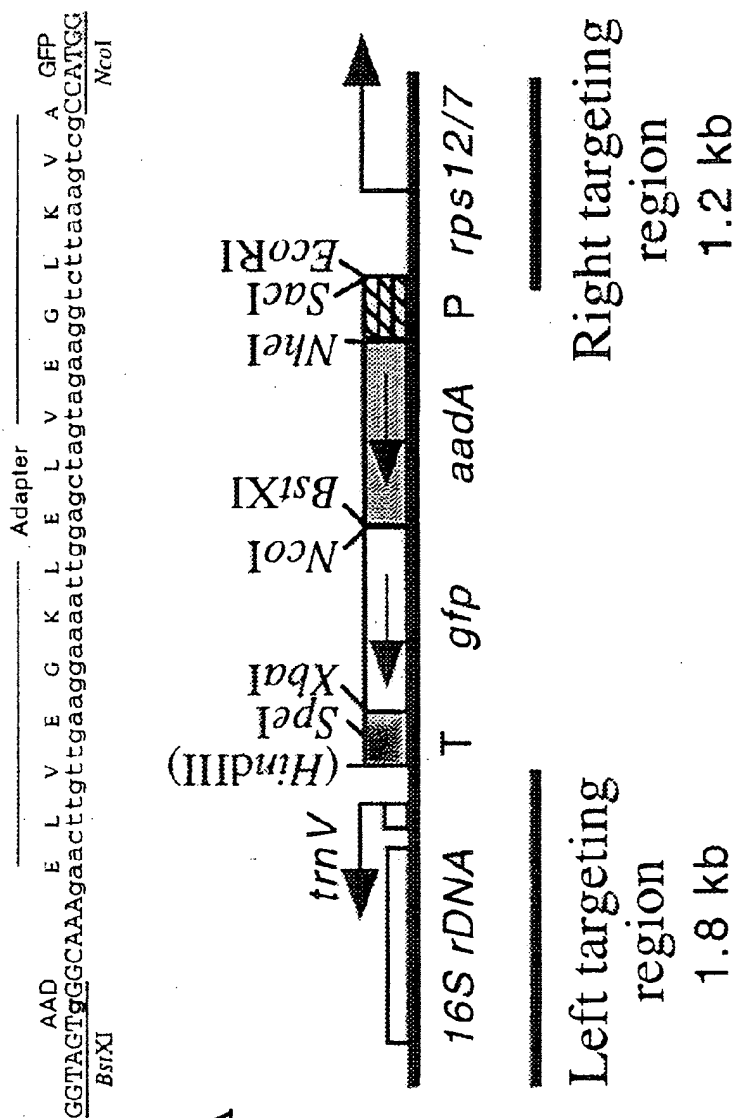


Figure 23A

Figure 23B

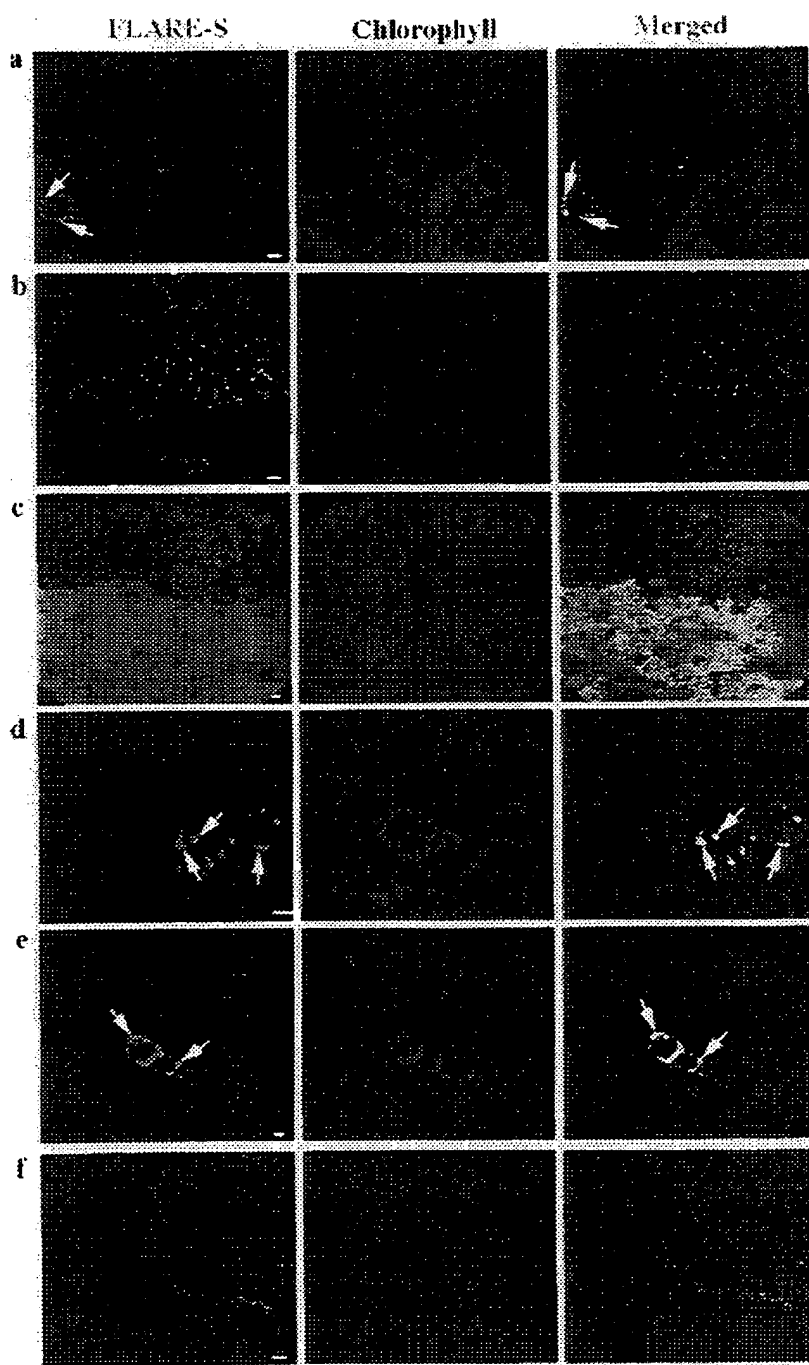


Figure 24

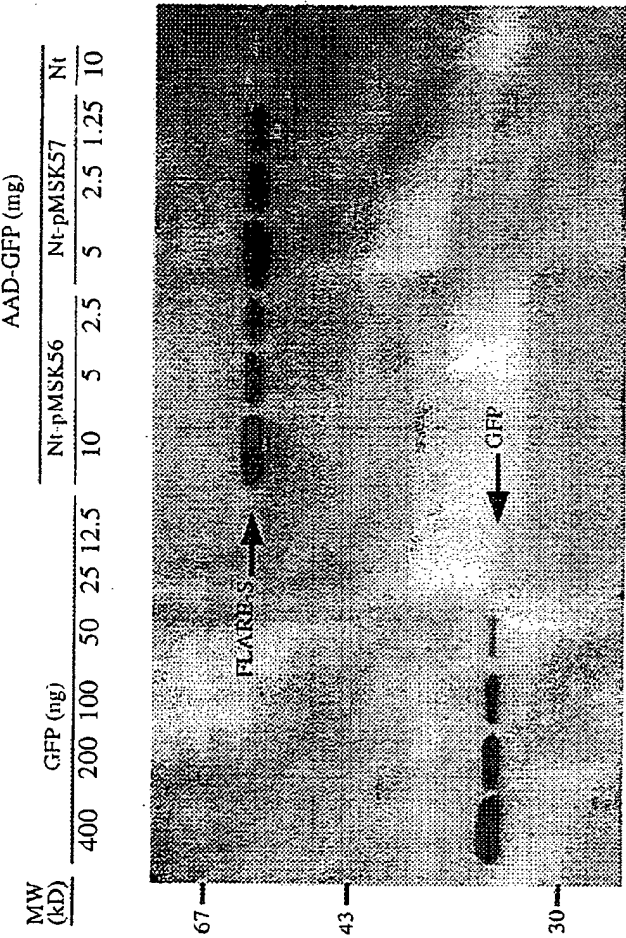


Figure 25

38/49

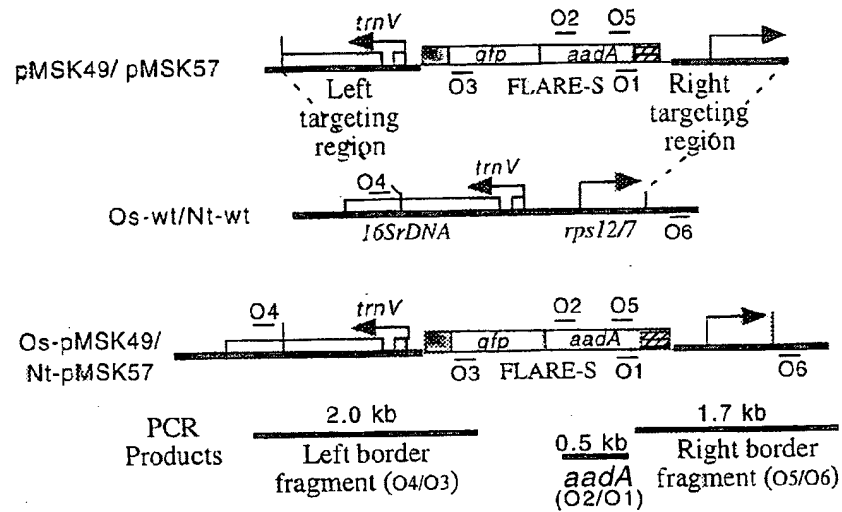


Figure 26A

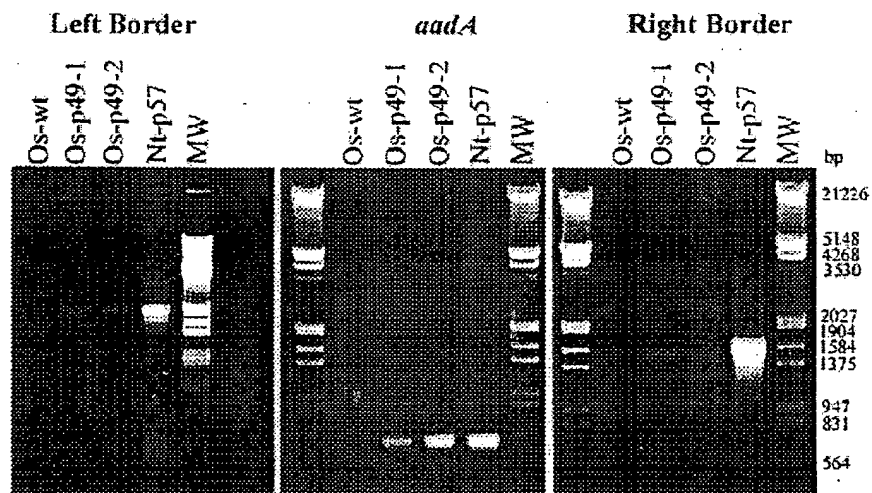


Figure 26B



Figure 27

FLARE16-S.seq Length: 1574

1 **NcoI** ccattgaggc tagcgaagcg gtgatcgccg aagtatcgac tcaactatca  
 51 GAGGTAGTTG GCGTCATCGA GCGCCATCTC GAACCGACGT TGCTGGCCGT  
 101 ACATTGTGAC GGCTCCSCAG TGGATGGCGG CCTGAAGCCA CACAGTGATA  
 151 TTGATTTGCT GGTACGGTG ACCGTAAAGTC TTGATGAAAC AACGCGGCGA  
 201 GCTTTGATCA ACGACCTTTT GGAAACTTCG GCTTCCCCTG GAGAGAGCGA  
 251 GATTCTCCGC GCTGTAGAAG TCACCATTGT TGTGCACGAC GACATCATTC  
 301 CGTGGCGTTA TCCAGCTAAG CGCGAATGTC AATTGGAGA ATGGCAGCGC  
 351 AATGACATTC TTGCAGGTAT CTTGAGGCCA GCCACGATCG ACATTGATCT  
 401 GGCTATCTTG CTGACAAAAG CAAGAGAACA TAGCGTTGCC TTGGTAGGTC  
 451 CAGCGGCGGA GGAACCTCTT GATCCGGTTC CTGAACAGGA TCTATTGAG  
 501 GCGCTAAATG AAACCTTAAC GCTATGGAAC TCGCCGCCCG ACTGGGCTGG  
 551 CGATGAGCGA AATGTAGTGC TTACGTTGTC CCGCATTGGG TACAGCGCAG  
 601 TAACCGGCAA AATCGCGCCG AAGGATGTCG CTGCCGACTG GGCAATGGAG  
 651 CGCCTGCCGG CCCAGTATCA GCGCGTCATA CTTGAAGCTA GACAGGCTTA  
 701 TCTTGACAA GAAGAAGATC GCTTGGCCTC GCGCGCAGAT CAGTTGGAAG  
 751 AATTGTCCA CTACGTGAAA GCGGAGATCA CCAAGGTAGT GGGCAAAgaa  
 801 **XbaI** ctgttggaag gaaaattgga gctagtagaa ggtcttaaag tcgccATGgc  
 851 EAGTAAAGGA GAAGAATTTT TCACTGGAGT TGTCCCAATT CTGTGTGAAT  
 901 TAGATGGTGA TGTTAATGGG CACAAATTTT CTGTCAGTGG AGAGGGTGAA  
 951 GGTGATGCAA CATACGGAAA ACTTACCTTT AAATTTATTT GCACTACTGG  
 1001 AAAACTACCT GTTCCTTGGC CAACACTTGT CACTACTTTC TCTTATGGTG  
 1051 TTCAATGCTT TTCAAGATAC CCAGATCATA TGAAGCGGCA CGACTTCTTC  
 1101 AAGAGCGCCA TGCTGAGGG ATACGTGCAG GAGAGGACCA TCTCTTTCAA  
 1151 GGACGACGGG AACTACAAGA CACGTGCTGA AGTCAAGTTT GAGGGAGACA  
 1201 CCTCGTCAA CAGGATCGAG CTTAAGGGAA TCGATTTCAG GGAGGACGGA  
 1251 AACATCCTCG GCCACAAGTT GGAATACAA TACAACCTCC ACAACGTATA  
 1301 CATCACGGCA GACAAACAAA AGAATGGAAT CAAAGCTAAC TTCAAATTA  
 1351 GACACAACAT TGAAGATGGA AGCGTTCAAC TAGCAGACCA TTATCAACAA  
 1401 AATACTCCAA TTGGCGATGG CCCTGTCCTT TTACCAGACA ACCATTACCT  
 1451 GTCCACACAA TCTGCCCTTT CGAAAGATCC CAACGAAAAG AGAGACCACA  
 1501 TGGTCCTTCT TGAGTTTGTG ACAGCTGCTG GATTACACA TGGCATGGAT  
 1551 GAACTATACA AATAAGctc taga

XbaI

Figure 28

FLARE16-S1.seq Length: 1953

1 **SacI**  
 gagctcgtc ccccgccgtc gttcaatgag aatggataag aggctcgtgg  
 51 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG TCGAGTAGAC  
 101 CTTGTTGTTG TGAAaAATTCT TAATTCATGA GTTGTAGGGA GGGATTTATG  
 151 TCACCACAAA CAGAGACTAA AGCAAGTGT GGATTCAAA ctagcTAAGC  
 201 GGTGATCGCC GAAGTATCGA CTCAACTATC AGAGGTAGTT GCGTCATCG  
 251 AGCGCCATCT CGAACCGACG TTGCTGGCCG TACATTTGTA CGGCTCCGCA  
 301 GTGGATGGCG GCCTGAAGCC ACACAGTGAT ATTGATTGTC TGTTACGGT  
 351 GACCGTAAGG CTTGATGAAA CAACCGCGCG AGCTTTGATC AACGACCTTT  
 401 TGGAAACTTC GGCTTCCCCT GGAGAGAGCG AGATTCTCCG CGCTGTAGAA  
 451 GTCACCATTG TTGTGCACGA CGACATCATT CCGTGGCGTT ATCCAGCTAA  
 501 GCGCGAACTG CAATTGGAG AATGGCAGCG CAATGACATT CTTCAGGTA  
 551 TCTTCGAGCC AGCCACGATC GACATTGATC TGGCTATCTT GCTGACAAA  
 601 GCAAGAGAAC ATAGCGTTGC CTGGTAGGT CCAGCGGCGG AGGAACTCTT  
 651 TGATCCGGTT CCTGAACAGG ATCTATTTGA GCGCTAAAT GAAACCTTAA  
 701 CGCTATGGA CTCGCGCGCC GACTGGGCTG GCGATGAGCG AAATGTAGTG  
 751 CTTACGTTGT CCGCATTTG GTACAGCGCA GTAACCGGCA AAATCGCGCC  
 801 GAAGGATGTC GCTGCCGACT GGGCAATGGA GCGCCTGCCG GCCCAGTATC  
 851 AGCCCGTCAT ACTTGAAGCT AGACAGGCTT ATCTTGGACA AGAAGAAGAT  
 901 CGCTTGGCCT CGCGCGCAGA TCAGTTGGAA GAATTTGTCC ACTACGTGAA  
 951 AGGCGAGATC ACCAAGGTAG TGGGCAAGa acttgttgaa ggaacattgg  
 1001 agctagtaga aggtcttaaa gtcgcccTgG ctAGTAAAGG AGAAGAACTT  
 1051 TTCACTGGAG TTGTCCCAAT TCTTGTGAA TTAGATGGTG ATGTTAATGG  
 1101 GCACAAATTT TCTGTCACTG GAGAGGGTGA AGGTGATGCA ACATACGGAA  
 1151 AACTTACCCT TAAATTTATT TGCACTACTG GAAACTACC TGTTCCTTGG  
 1201 CCAACACTTG TCACTACTTT CTCTTATGGT GTTCAATGCT TTTCAGATA  
 1251 CCCAGATCAT ATGAAGCGC ACGACTTCTT CAAGAGCGCC ATGCCTGAGG  
 1301 GATACGTGCA GGAGAGGACC ATCTCTTCA AGGACGACGG GAACACAAAG  
 1351 ACACGTGCTG AAGTCAAGTT TGAGGGAGAC ACCCTCGTCA ACAGGATCGA  
 1401 GCTTAAGGGA ATCGATTTC AGGAGGACGG AAACATCCTC GGCCACAAGT  
 1451 TGGAATACAA CTACAACTCC CACAACGTAT ACATCACGGC AGACAAACAA  
 1501 AAGAATGGAA TCAAAGCTAA CTTCAAATTT AGACACAACA TTGAAGATGG  
 1551 AAGCGTTCAA CTAGCAGACC ATTATCAACA AAATACTCCA ATTGGCGATG  
 1601 GCCCTGTCTT TTTACCAGAC AACCATTACC TGTCCACACA ATCTGCCCTT  
 1651 TCGAAAGATC CCAACGAAA GAGAGACCAC ATGGTCCTTC TTGAGTTTGT  
 1701 AACAGCTGCT GGGATTACAC ATGGCATGGA TGAACATAC AAATAAGct  
 1751 ctagagccAT CCTGGCTTAG TCTATAGGAG GTTTTGAAAA GAAAGGAGCA  
 1801 ATAATCATT TCTTGTCTA TCAAGAGGCT GCTATTGCTC CTTTCTTTT  
 1851 TTCTTTTAT TTATTTACT GTATTTTACT TACATAGACT TTTTGTTTA  
 1901 CATTATAGAA AAAGAAGGAG AGGTATTTT CTTCATTTA TTCAATGaaag  
 1951 ctt

Lotp8 28  
 oada  
 jfp  
 Tps6A  
 Hind III

Figure 29



FLARE16-S2.seq Length: 1985

1 *SacI*  
 51 gagctcgc CTC CCCCGCCGTC GTTCAATGAG AATGGATAAG AGGCTCGTGG  
 101 GATTGACGTG AGGGGGCAGG GATGGCTATA TTTCTGGGAG AATTAACCGA  
 151 TCGACGTGCa AGCGGACATT TATTTTaaAT TCGATAATTT TTGCAAAAC  
 201 ATTTTCGACAT ATTTATTTAT TTTATTATTA TGAGAATCAA TCCTACTACT  
 251 TCTGGTTCIG GGGTTTCCAC GgctagcGAA GCGGTGATCG CCGAAGTATC  
 301 GACTCAACTA TCAGAGGTAG TTGGCGTCAT CGAGCGCCAT CTCGAACCGA  
 351 CGTTGCTGGC CGTACATTIG TACGGCTCCG CAGTGGATGG CGGCTGAAG  
 401 CCACACAGTG ATATTGATT TCTGGTTACG GTGACCGTAA GGCTTGATGA  
 451 AACACGCGG CGAGCTTTGA TCAACGACCT TTTGGAACT TCGGCTTCCC  
 501 TGGTACAGCG CAGTAACCGG CAAATCGCG CCGAAGGATG TCGCTGCCGA  
 551 GACGACATCA TTCCGTGGCG TTATCCAGCT AAGCGCGAAC TGCAATTGGG  
 601 AGAATGGCAG CGCAATGACA TTCTTGACGG TATCTTCGAG CCAGCCACGA  
 651 TCGACATTGA TCTGGCTATC TTGCTGACAA AAGCAAGAGA ACATAGCGTT  
 701 GCCTTGGTAG GTCCAGCGGC GGAGGAACTC TTTGATCCGG TTCCTGAACA  
 751 GGATCTATTT GAGGCGCTAA ATGAAACCTT AACGCTATGG AACTCGCCGC  
 801 CCGACTGGGC TGGCGATGAG CGAAATGTAG TGCTTACGTT GTCCCGCATT  
 851 TGGTACAGCG CAGTAACCGG CAAATCGCG CCGAAGGATG TCGCTGCCGA  
 901 CTGGGCAATG GAGCGCCTGC CGGCCCGTA TCAGCCCGTC ATACTTGAAG  
 951 CTAGACAGGC TTATCTTGGG CAAGAAGAAG ATCGCTTGGC CTCGCGCGCA  
 1001 GATCAGTTGG AAGAATTGT CCACTACGTG AAAGGCGAGA TCACCAAGGT  
 1051 AGTGGGCAAA gaacttggtg aaggaaaatt ggagctagta gaaggtctta  
 1101 aagtcgccAT GgctAGTAAA GGAGAAGAAC TTTTCACTGG AGTTGTCCCA  
 1151 ATTCTTGTG AATTAGATGG TGATGTTAAT GGGCACAAT TTTCTGTCAG  
 1201 TGGAGAGGGT GAAGGTGATG CAACATACGG AAAACTTACC CTTAAATTTA  
 1251 TTTGCACTAC TGGAAACTA CTTGTTCCTT GGCCAACACT TGTCACTACT  
 1301 TTCTCTTATG GTGTTCAATG CTTTTCAAGA TACCCAGATC ATATGAAGCG  
 1351 GCACGACTTC TTCAAGAGCG CCATGCCTGA GGGATACGTG CAGGAGAGGA  
 1401 CCATCTCTTT CAAGGACGAC GGGAACTACA AGACACGTGC TGAAGTCAAG  
 1451 TTTGAGGGAG ACACCTTCGT CAACAGGATC GAGCTTAAGG GAATCGATT  
 1501 CAAGGAGGAC GGAAACATCC TCGGCCACAA GTTGGAAATC AACTACAAT  
 1551 CCCACAACGT ATACATCAGG GCAGACAAAC AAAAGAATGG AATCAAAGCT  
 1601 AACTTCAAAA TTAGACACAA CATTGAAGAT GGAAGCGTTC AACTAGCAGA  
 1651 CCATTATCAA CAAAATACTC CAATGGCGA TGGCCCTGTC CTTTTACCAG  
 1701 ACAACCATTA COTGTCCACA CAATCTGCC TTTGAAAGA TCCCAACGAA  
 1751 AAGAGAGACC ACATGGTCCT TCTTGAGTTT GTAACAGCTG CTGGGATTAC  
 1801 ACATGGCATG GATGAATAT ACABAATAAGg ctctagagc ATCCTGGCCT  
 1851 AGTCTATAGG AEGTTTGA AAGAAAGGAG CAATAATCAT TTTCTGTTC  
 1901 TATCAAGAGG GTGCTATTGC TCCTTCTTT TTTCTTTT ATTTATTTAC  
 1951 TAGTATTTTA CTTACATAGA CTTTTTGT TACATTATAG AAAAGAAGG  
 AGAGGTTATT TTCTTGCAAT TATTCATGaa agctt  
 HindIII

*Labels:*  
 L76LDB  
 oadA  
 gfp  
 TpsbA

Figure 30

FLARE11-S.seq Length: 1595

NcoI                      C-Myc

```

1  ccattgggggc tagcgaacaa aaactcattt ctgaagaaga cttgcctagc
51  GAAGCGGTGA TCGCCGAAGT ATCGACTCAA CTATCAGAGG TAGTTGGCGT
101 CATCGAGCGC CATCTCGAAC CGACGTTGCT GGCCGTACAT TTGTACGGCT
151 CCGCAGTGGI. TGGCGGCCTG AAGCCACACA GTGATATTGA TTTGCTGCTT
201 ACGGTGACCG TAAGGCTTGA TGAACAACG CGGCGAGCTT TGATCAACGA
251 CCTTTTGGAA ACTTCGGCTT CCCCTGGAGA GAGCGAGATT CTCCGCGCTG
301 TAGAAGTCAC CATTGTGTG CACGACGACA TCATTCCGTG GCGTTATCCA
351 GCTAAGCGCG AACTGCAATT TGGAGAATGG CAGCGCAATG ACATTCTTGC
401 AGGTATCTTC GAGCCAGCCA CGATCGACAT TGATCTGGCT ATCTTGCTGA
451 CAAAAGCAAG AGAACATAGC GTTGCCCTTG TAGGTCCAGC GGCGGAGGAA
501 CTCTTTGATC CGGTTCTTGA ACAGGATCTA TTTGAGGCGC TAAATGAARC
551 CTTAACGCTA TGGAACTCGC CGCCCGACTG GGCTGGCGAT GAGCGAAATG
601 TAGTGCTTAC GTTGTCCTGC ATTTGGTACA GCGCAGTAAC CGGCAAAATC
651 GCGCCGAAGG ATGTCGCTGC CGACTGGSCA ATGGAGCGCC TGCCGGCCCA
701 GTATCAGCCC GTCATACTTG AAGCTAGACA GGCTTATCTT GGACAAGAAG
751 AAGATCGCTT GGCCTCGCGC GCAGATCAGT TGGAAAGAATT TGTCCACTAC
801 GTGAAGGCGG AGATCACCAA GGTAGTGGC AAAaacttg cagttgaagg
851 aaaattggag gtcgcccTGG ctAGTAAAGG AGAAGAACTT TTCACTGGAG
901 TTGTCCCAAT TCTTGTGAA TTAGATGGTG ATGTTAATGG GCACAAATTT
951 TCTGTCAGTG GAGAGGGTGA AGGTGATGCA ACATACGGAA AACTTACCTT
1001 TAAATTTATT TGCACTACTG GAAAACTACC TGTTCCTTGG CCAACRCTTG
1051 TCACCTACTT CTCTTATGGT GTTCAATGCT TTTCAAGATA CCCAGATCAT
1101 ATGAAGCGGC ACGACTTCTT CAAGAGCGCC ATGCCTGAGG GATACGTGCA
1151 GGAGAGGACC ATCTCTTTCA AGGACGACGG GAACTACAAG ACACGTGCTG
1201 AAGTCAAGTT TGAGGGAGAC ACCCTCGTCA ACAGGATCGA GCTTAAGGGA
1251 ATCGATTTCA AGGAGGACGG AACATCCTC GGCCACAAGT TGGAAATACAA
1301 CTACAACCTC CACAACGTAT ACATACGGC AGACAAACAA AAGAATGGAA
1351 TCAAAGCTAA CTTCAAAATT AGACACAACA TTGAAGATGG AAGCGTTCAA
1401 CTAGCAGACC ATTATCAACA AAATACTCCA ATTGGCGATG GCCCTGTCTT
1451 TTTACCAGAC AACCATTACC TGTCCACACA ATCTGCCCTT TCGAAAGATC
1501 CCAACGAAAA GAGAGACCAC ATGGTCCTTC TTGAGTTTGT AACAGCTGCT
1551 GGGATTACAC ATGGCATGGA TGAATATAC AAATAAGgct ctaga

```

aada  
dSP  
XbaI

Figure 31

FLARE11-S3.seq Length: 1961

1 Salt  
 51 gaactcgcctc ccccgccgctc gttcaatgag aatggataag aggctcgtgg  
 101 gattgacgtg agggggcagg gatggctata tttctgggag ggagaccaca  
 151 acggtttccc actagaaata atTTTgTTTA acTTTAAGAA ggagatatac  
 201 BTATGGCaAG CATGACTGGT GGACAGgcta gCGaacaanaa actcatttct  
 251 gaagaagact tgcctagc GA AGCGGTGATC GCCGAAGTAT CGACTCAACT  
 301 ATCAGAGGTA GTTGGCGTCA TCGAGCGCCA TCTCGAACCG ACGTTGCTGG  
 351 CCGTACATTT GTACGGCTCC GCACTGGATG GCGGCCTGAA GCCACACAGT  
 401 GATATTGATT TGCTGGTTAC GGTGACCGTA AGGCTTGATG AAACAACGCG  
 451 GCGAGCTTTG ATCAACGACC TTTTGGAAAC TTGCGCTTCC CCTGGAGAGA  
 501 GCGAGATTCT CCGCGCTGTA GAAGTCACCA TTGTTGTGCA CGACGACATC  
 551 ATTCCGTGGC GTTATCCAGC TAAGCGCGAA CTGCAATTTG GAGAATGGCA  
 601 GCGCAATGAC ATTCTTGACG GTATCTTCGA GCCAGCCACG ATCGACATTG  
 651 ATCTGGCTAT CTTGCTGACA AAAGCAAGAG AACATAGCGT TGCTTTGGTA  
 701 GGTCCAGCGG CGGAGGAACT CTTTGATCCG GTTCCTGAAC AGGATCTATT  
 751 TGAGGCGCTA AATGAAACCT TAACGCTATG GAACTCGCCG CCCGACTGGG  
 801 CTGGCGATGA GCGAAATGTA GTGCTTACGT TGTCCCGCAT TTGGTACAGC  
 851 GCAGTAACCG GCAAAATCGC GCCGAAGGAT GTCGCTGCCG ACTGGGCAAT  
 901 GGAGCGCCTG CCGGCCCACT ATCAGCCCGT CATACTTGAA GCTAGACAGG  
 951 CTTATCTTGG ACAAGAAGAA GATCGCTTGG CCTCGCGCGC AGATCAGTTG  
 1001 GAAGAATTTG TCCACTACGT GAAAGGCGAG ATCACCBAAG TAGTGGGCAA  
 1051 gaacttgca gttgaaggaa aattggaggt cgccATGgct AGTAAAGGAG  
 1101 AAGAATTTT CACTGGAGTT GTCCCAATTC TTGTTGAATT AGATGGTGAT  
 1151 GTTAATGGGC ACAAATTTTC TGTCAGTGGA GAGGGTGAA GTGATGCAAC  
 1201 ATACGGAAAA CTTACCCTTA AATTTATTTG CACTACTGGA AACTACCTG  
 1251 TTCCtTGGCC AACACTTGTC ACTACTTTCT CTTATGGTGT TCAATGCTTT  
 1301 TCAAGATAAC CAGATCATAT GAAGCGGCAC GACTTCTTCA AGAGCGCCAT  
 1351 GCCTGAGGGA TACGTGCAGG AGAGGACCAT CTCTTTCAAG GACGACGGGA  
 1401 ACTACAAGAC ACGTGCTGAA GTCAAGTTTG AGGGAGACAC CCTCGTCAAC  
 1451 AGGATCGAGC TTAAGGGAAT CGATTTCAG GAGGACGGAA ACATCCTCGG  
 1501 CCACAAGTTG GAATACAAC ACAACTCCCA CAACGTATAC ATCACGGCAG  
 1551 ACAAACAAAA GAATGGAATC AAAGCTAACT TCAAAATTAG ACACAACATT  
 1601 GAAGATGGAA GCGTTCAACT AGCAGACCAT TATCAACAAA ATACTCCAAT  
 1651 TGGCGATGGC CCTGTCTTTT TACCAGACAA CCATTACCTG TCCACACAAT  
 1701 CTGCCCTTTC GAAAGATCCC AACGAAAAGA GAGACCACAT GGTCCTTCTT  
 1751 GAGTTTGTAA CAGCTGCTGG GATTACACAT GGCATGGATG AACTATACAA  
 1801 ATAAGgctct agagcCATCC TGGCCTAGTC TATAGGAGGT TTTGAAAAGA  
 1851 AAGGAGCAAT AATCATTTTC TTGTTCTATC AAGAGGGTGC TATTGCTCCT  
 1901 TTCTTTTTTT CTTTTTATTT ATTTACTAGT ATTTTACTTA CATAGACTTT  
 1951 TTTGTTTTACA TTATAGAAAA AGAAGGAGAG GTTATTTTCT TGCATTTATT  
 CATGaaagct t  
 HindIII

Figure 32

pMSK35.seq Length: 4671

## Figure 33A

122488

1	GGGAACGGAT	TCACCGCCGT	ATGGCTGACC	GGCGATTACT	AGCGATTCTT
51	GCTTCATGCA	GGCGAGTTGC	AGCCTGCAAT	CCGAACTGAG	GACGGGTTTT
101	TGGAGTTAGC	TCACCTTCGC	GAGATCGCGA	CCCTTTGTCC	CGCCCATTTG
151	AGCACGTGTG	TCGCCCAGGG	CATAAGGGGC	ATGATGACTT	GGCCTCATCC
201	TCTCCTTCCT	CCGGCTTAAC	ACCGGCGGTC	TGTTTCAGGGT	TCCAAACTCA
251	TAGTGGCAAC	TAAACACGAG	GGTTGCGCTC	GTTGCGAGAC	TTAACCCAAC
301	ACCTTACGGC	ACGAGCTGAC	GACAGCCATG	CACCACCTGT	GTCCGGGTTT
351	CCGAGGGCAC	CCCTCTCTTT	CAAGAGGATT	CGCGGCATGT	CAAGCCCTGG
401	TAAGGTTCTT	CGCTTTGCAT	CGAATTAAAC	CACATGCTCC	ACCGCTTGTG
451	CGGGCCCCCG	TCAATTCCTT	TGAGTTTCAT	ICTTGCGAAC	GTACTCCCCA
501	GGCGGGATAC	TTAACGCGTT	AGCTACAGCA	CTGCACGGGT	CGAGTCGCAC
551	AGCACCTAGT	ATCCATCGTT	TACGGCTAGG	ACTACTGGGG	TCTCIAATCC
601	CATTTGCTCC	CCTAGCTTTC	GTCTCTCAGT	GTCACTGTTC	GCCGAGCAGA
651	GTGCTTTTCG	CGTTGGTGTT	CTTTCCGATC	TCAATGCATT	TCACCGCTCC
701	ACCGGAAATT	CCCTCTGCCC	CTACCGTACT	CCAGCTTGGT	AGTTTCCACC
751	GCCTGTCCAG	GGTTGAGCCC	TGGGATTGTA	CGGCGGACTT	GAAAAGCCAC
801	CTACAGACGC	TTTACGCCCC	ATCATTCCGG	ATAACGCTTG	CATCCTCTGT
851	CTTACCGCGG	CTGCTGGCAC	AGAGTTAGCC	GATGCTTATT	CCTCAGATAC
901	CGTCATTGTT	TCTTCTCGA	GAAAAGAAGT	TGACGACCCG	TGGGCCTTCC
951	ACCTCCACGC	GGCATTGCTC	CGTCAGGCTT	TCGCCCATTG	CGGAAATTC
1001	CCCCTGTCTG	CCTCCGCTAG	GAGTCTGGGC	CGTGTCTCAG	TCCCAGTGTG
1051	GCTGATCATC	CTCTCGGACC	AGCTACTGAT	CATCGCCTTG	GTAAGCTATT
1101	GCCTCACCAA	CTAGCTAATC	AGACGCGAGC	CCCTCCTTGG	GCGGATTTC
1151	CCTTTGTCTC	CTCAGCCTAC	GGGGTATTAG	CAACCGTTTC	CAGTTGTGTG
1201	TCCCCTCCCA	AGGGCAGGTT	CTTACGCGTT	ACTCACCCGT	TCGCCACTGG
1251	AAACACCACT	TCCCGTTCGA	CTTGCACTGT	TTAAGCATGC	CGCCAGCGTT
1301	CATCCTGAGC	CAGGATCGAA	CTCTCCATGA	GATTCATAGT	TGCATTACTT
1351	ATAGCTTCCT	TATTCGTAGA	CAAAGCGGAT	TCCGAATTGT	CTTTCCTTCC
1401	AAGGATAACT	TGTATCCATG	CGCTTCAGAT	TATTAGCCTG	GAGTTCGCCA
1451	CCAGCAGTAT	AGCCAACCCT	ACCTATCAC	GTCAATCCCA	CAAGCCTCTT
1501	ATCCATTCCC	GTTGATCGT	GGCGGGGGGA	GTAAGTCAAA	ATAGAAAAAA
1551	CTCACATTGG	GTTTAGGGAT	AATCAGGCTC	GAACTGATGA	CTTCCACCAC
1601	GTCAAGGTGA	CACCTACCG	CTGAGTTATA	TCCCTTCCGC	GTCCCTCGA
1651	GAAAGAGAAT	TACCGAATCC	TAAGGCAAG	GGGCGAGAAA	CTCAAGGCCA
1701	CCCTTCTCTC	GGGCTTTCTT	TCCACACTAT	TATGGATAGT	CAAATAATGG
1751	GAAAAATTGG	ATTCAATTGT	CAACCGGTCC	TATCGAAAAA	AGGATTGACT
1801	ATGGATTGCA	GCCATAGCAC	ATGGTTTCAT	AAAATCTGTA	CGATTTTCCC
1851	GATCTAATC	GAGCAGGTTT	CCATGAAGAT	gatcgacggt	atcgataagc
1901	ttgcatgcct	gcaggtCGAA	TATAGCTCTT	CTTCTTATT	TCAATGATAT
1951	TATTATTTCA	AAGATAAGAG	ATATTCAAAG	ATAAGAGATA	AGAAGAAGTC
2001	AAAATTTGAT	TTTTTTTTTG	GAAAAAATA	ATCAAAAAGA	TATAGTAACA
2051	TTAGCAAGAA	GAGAAACAAG	TTCTATTTCA	CAATTTAAAC	AAATACAAAA
2101	TCAAAATAGA	ATACTCAATC	ATGAATAAAT	GCAAGAAAAA	AACCTCTCCT
2151	TCTTTTTCTA	TAATGTAAAC	AAAAAAGTCT	ATGTAAGTAA	AATACTAGTA
2201	AATAAATAAA	AAGAAAAAAA	GAAAGGAGCA	ATAGCACCCT	CTTGATAGAA
2251	CAAGAAATG	ATTATTGCTC	CTTCTTTTTC	AAAACCTCCT	ATAGACTAGG
2301	CCAGGATCgc	tctagcTAGA	CATTATTTGC	CGACTACCTT	GTTGATCTCG
2351	CCTTTTACGT	AGTGGACAAA	TTCTTCCAA	TGATCTGCGC	GCGAGGCCAA
2401	GCGATCTTCT	TCTTGTCCAA	GATAAGCCTG	TCTAGCTTCA	AGTATGACGG
2451	GCTGATACGT	GGCCGGCAGG	CGCTCCATTG	CCCAGTCGGC	AGCGACATCC
2501	TTCCGGCGCA	TTTTGCCGGT	TACTGCGCTG	TACCAATGCG	GGGACAACGT
2551	AAGCACTACA	TTTCGCTCAT	CGCCAGCCCA	GTCCGGCGGC	GAGTTCCATA
2601	GCGTTAAGGT	TTCAATTAGC	GCCTCAAATA	GATCCTGTTT	AGGAACCGGA
2651	TCAAAAGAGT	CCTCCGCCGC	TGGACCTACC	AAGGCAACGC	TATGTTCTCT
2701	TGCTTTTGTG	AGCAAGATAG	CCAGATCAAT	GTCGATCGTG	GCTGGCTCGA

Rice left targeting sequence

cadda

pMSK35.seq Length: 4671

```

2751 AGATACCTGC AAGAATGTCA TTGCGCTGCC ATTCTCCAAA TTGCAGTTCC
2801 CGCTTAGCTG GATAACGCCA CGGAATGATG TCGTCGTGCA CAACATGGT
2851 GACTTCTACA GCGCGGAGAA TCTCGCTCTC TCCAGGGGAA GCCGAAGTTT
2901 CCAAAGGTC GTTGATCAAA GCTCGCGCGG TTGTTTCATC AAGCCTTACG
2951 GTCACCGTAA CCAGCAAATC AATATCACTG TGTGGCTTCA GGCCGCCATC
3001 CACTGCGGAG CCGTACAAAT GTACGCCAG CAACGTCGGT TCGAGATGGC
3051 GCTCGATGAC GCCAACTACC TCTGATAGTT GAGTCGATAC TTCGGCGATC
3101 ACCGCTTCCC TCATGgATCC CTCCCTACAA CTGTATCCAA GCGCTTCgTA
3151 TTCGCCCCGA GTTCGCTCCC AGAAATATAG CCATCCCTGC CCCCTCACGT
3201 CAATCCCACG AGCCTCTTAT CCATTCTCAT TGAACGACGG CGGGGGAGC
3251 ttgggtaccg agctcgaatt cctgcagccc gatcTTACCA TTTCGGAAGG
3301 AACTGGGGCT ACATTTCTTT TCAATTTCCA TTCAAGAGTT TCTTATCTGT
3351 TTCCACGCCC TTTTITGAGA CCTCGAAACA TGAAATGGAC AAATTCCTTC
3401 TCTTAGGAAC ACATACAAGA AAAAGGATAA TGGTAGCCCT CCCATTAAct
3451 ACTTCATTTT ATTTATGAAT TTCATAGTAA TAGAAATCCA TGTCCTACCG
3501 AGACAGAATT TCGAACTTGC TATCCTCTTG CCTAATAGGC AAAGATTGAC
3551 CTCTGTAGAA AGAATGATTC ATTCCGATCG ATATGAGGAC CCAACTACGT
3601 TCGATTGCAG AATCCATGTT CCATATTTGA AGAGGGTTGA CCTCTGTGCT
3651 TCTCTCATGG TACAATCCTC TTCTGTCTGA GCCCCCTTTC TCCTCGGTCC
3701 ACAGAGAAAA AATGGAGGAC TGGTGCAGAC AGTTCATCAC GGAAGAAAGA
3751 ACTCACAGAG CCGGGATCGC TAACTAATAG AATAGTACTA CTAATAATA
3801 CTAATATATA GAAATAGATA Tctagctaga AATAGAAACA ACTAATATAT
3851 AGATAATCGA AATTGAAAAG AACTGTCTTT TCTGTATACT TTCCCCGTTT
3901 TATTGCTACC GCGGGTCTTA TGCAATCGAT CGGATCATAT AGATATCCCT
3951 TCAACACAAC ATAGGTCATC GAAAGGATCT CGGACGACTC ACCAAAGCAC
4001 GAAAGCCAGT TAGAAAATGG ATTCCTATTT GAAGAGTGCC TAACCGCATG
4051 GATAAGCTCA CATTAAACCG TCAATTTTGG ATCCAATTCT GGATTTTCTT
4101 TGGGAAGTTT CGGGAAGAAA TTGGAATGGA ATAATATAGA TTCATACAGA
4151 GGAAAAGGTT CTCTATTGAT GCAAACGCTG TACCTAGAGG ATAGGGATAG
4201 AGGAAGAGGG AAAAATCGAA ATGAAATAAA TAAGAATAAA AGCAAAAAAA
4251 AAATAAGTCG AAGATAGAG AGCCAGATT CCAATGAAG AAATGGAAAC
4301 TCGAAAAGGA TCCTTCTGAT TCTCAAAGAA TGAGGGGCAA GGGGATTGAT
4351 ACCGAGAAAG ATTTCTTCTT ATTATAAGAC GTGATTTGAT CCGCATATGT
4401 TTGGTAAAAG AACAATCTTC TCCTTTAATC ATAAATGGAA AGTGTTCAAT
4451 TAGAACATGA AAACGTGACT CAATTGGTCT TAGTTAGTCT TCGGGACGGA
4501 GTGGAAGAAA GGGCGAAGAC TCTCGAACGA GGAAAAGGAT CCCTTCGAAA
4551 GAATTGAACG AGGAGCCGTA TTAGGTGAAA ATCTCATGTA CGATTCTGTA
4601 GAGGGACAGG AAGGGTGAAT TATCTGTGGA CTTTCCACT ATCAACCCCA
4651 AAAAACCCTA CTCTGCCTTA C

```

*add*  
*Rice Right-targeting sequence*

125876

Figure 33B

pMSK49.seq Length: 5263

Figure 34A

122488

1	GGGAACGGAT	TCACCGCCGT	ATGGCTGACC	GGCGATTACT	AGCGATTCTT
51	GCTTCATGCA	GGCGAGTTGC	AGCCTGCAAT	CCGAACGTAG	GACGGGTTTT
101	TGGAGTTAGC	TCACCCTCGC	GAGATCGCGA	CCCTTTGTCC	CGCCCATGTG
151	AGCACGTGTG	TCGCCCAGGG	CATAAGGGGC	ATGATGACTT	GGCCTCATCC
201	TCTCCTTCCT	CCGGCTTAAC	ACCGGCGGTC	TGTTCAGGGT	TCCAAACTCA
251	TAGTGGCAAC	TAAACACGAG	GGTTGCGCTC	GTTGCGAGAC	TTAACCCAAC
301	ACCTTACGGC	ACGAGCTGAC	GACAGCCATG	CACCACCTGT	GTCCGCGTTC
351	CCGAGGGCAC	CCCTCTCTTT	CAAGAGGATT	CGCGGCATGT	CAAGCCCTGG
401	TAAGGTTCTT	CGCTTTGCA	CGAATTAAAC	CACATGCTCC	ACCGCTTGTC
451	CGGGCCCCCG	TCAATTCCTT	TCAGTTTCAT	TCTTGCGAAC	GTACTCCCCA
501	GGCGGGATAC	TTRACGCGTT	AGCTACAGCA	CTGCACGGGT	CGAGTCGCAC
551	AGCACCTAGT	ATCCATCGTT	TACGGCTAGG	ACTACTGGGG	TCTCTAATCC
601	CATTTGCTCC	CCTAGCTTTC	GTCTCTCAGT	GTCAGTGTCG	GCCCAGCAGA
651	GTGCTTTCGC	CGTTGGTGT	CTTTCGGATC	TCAATGCATT	TCACCGCTCC
701	ACCGGAAATT	CCCTCTGCCC	CTACCGTACT	CCAGCTTGGT	AGTTTCCACC
751	GCCTGTCCAG	GGTTGAGCCC	TGGGATTGTA	CGGCGGACTT	GAAAAGCCAC
801	CTACAGACGC	TTTACGCCCC	ATCATTCCGG	ATAACGCTTG	CATCCTCTGT
851	CTTACCGCGG	CTGCTGGCAC	AGAGTTAGCC	GATGCTTATT	CCTCAGATAC
901	CGTCATTGTT	TCCTCTCCGA	GAAAAGAAGT	TGACGACCCG	TGGGCCCTTC
951	ACCTCCACGC	GGCATTGCTC	CGTCAGGCTT	TCGCCCATTG	CGGAAATTC
1001	CCCCTGCTG	OCTCCCGTAG	GAGTCTGGGC	CGTGTCTCAG	TCCCAGTGTG
1051	GCTGATCATC	CTCTCGGACC	AGCTACTGAT	CATCGCCTTG	GTAAGCTATT
1101	GCCTCACCAA	CTAGCTAATC	AGACGCGAGC	CCCTCCTTGG	GCGGATTCTT
1151	CCCTTTGCTC	CTCAGCCTAC	GGGGTATTAG	CAACCGTTTC	CAGTTGTGTG
1201	TCCCTCCCA	AGGGCAGGTT	CTTACGCGTT	ACTCACCGGT	TCGCCACTGG
1251	AAACACCACT	TCCCCTTCGA	CTTGCAATGT	TTAAGCATGC	CGCCAGCGTT
1301	CATCTGAGC	CAGGATCGAA	CTCTCCATGA	GATTCATAGT	TGCATTACTT
1351	ATAGCTTCCT	TATTCCTAGA	CAAAGCGGAT	TCGGAATTGT	CTTTCCTTCC
1401	AAGGATAACT	TGTATCCATG	CGCTTCAGAT	TATTAGCCTG	GAGTTCGCCA
1451	CCAGCAGTAT	AGCCAAACCT	ACCCTATCAC	GTCAATCCCA	CAAGCCTCTT
1501	ATCCATTCCC	GTTCGATCGT	GGCGGGGGGA	GTAAGTCAAA	ATAGAAAAAA
1551	CTCACATTGG	GTTTAGGGAT	AATCAGGCTC	GAACTGATGA	CTTCCACCAC
1601	GTCAAGGTGA	CACCTTACCG	CTGAGTTATA	TCCCTTCCCC	GTCCCTTCGA
1651	GAAAGAGAAT	TACCGAATCC	TAAGGCAAG	GGGCGAGAAA	CTCAAGGCCA
1701	CCCTTCCTCC	GGGCTTCTT	TCCACACTAT	TATGGATAGT	CAAATAATGG
1751	GAAAAATTGG	ATTCAATTGT	CAACCGGTCC	TATCGAAAAT	AGGATTGACT
1801	ATGGATTGGA	GCCATAGCAC	ATGGTTTCAT	AAAATCTGTA	CGATTTCCTC
1851	GATCTAAATC	GAGCAGGTTT	CCATGAAGAA	gatcgacggt	atcgataagc
1901	TTTCTATGAT	AAATGCAAGA	AAATAACCTC	TCCTTCTTTT	TCTATATAGT
1951	AAACAAAAAA	GTCTATGTAA	GTAAATAACT	AGTAAATAAA	TAAAAAGAAA
2001	AAAGAAAGG	AGCAATAGCA	CCCTCTTGAT	AGAACAAGAA	AATGATTATT
2051	GCTCCTTCTT	TTTCAAAACC	TCCTATAGAC	TAGGCCAGGA	TCgctctaga
2101	gccttatttg	TATAGTTTCAT	CCATGCCATG	TGTAATCCCA	GCAGCTGTTA
2151	CAAACTCAAG	AAGGACCATG	TGGTCTCTCT	TTTCGTTGGG	ATCTTTTCGAA
2201	AGGGCAGATT	GTGTGGACAG	GTAATGGTTG	TCTGGTAAAA	GGACAGGGCC
2251	ATCGCCAATT	GGAGTATTTT	GTTGATAATG	GTCTGCTAGT	TGAACGCTTC
2301	CATCTTCAAT	GTTGTGTCTA	ATTTTGAAGT	TAGCTTTGAT	TCCATTCTTT
2351	TGTTTGTCTG	CCGTGATGTA	TACGTTGTGG	GAGTTGTAGT	TGTAATCCAA
2401	CTTGTGGCCG	AGGATGTTTC	CGTCCTCCTT	GAAATCGATT	CCCTTAAGCT
2451	CGATCCTGTT	GACGAGGGTG	TCTCCCTCAA	ACTTGACTTC	AGCACGTGTC
2501	TTGTAGTTCC	CGTCGTCTTT	GAAAGAGATG	GTCTCTCCTT	GCACGTATCC
2551	CTCAGGCATG	GCGCTCTTGA	AGAAGTCGTG	CCGCTTCATA	TGATCTGGGT
2601	ATCTTGAAAA	GCATTGAACA	CCATAAGAGA	AAGTAGTGAC	AAGTGTGGCC
2651	CAAGGAACAG	GTAGTTTTC	AGTAGTGCAA	ATAAATTPTAA	GGGTAAGTTT
2701	TCCGTATGTT	GCATCACCTT	CACCTCTCTC	ACTGACAGAA	AATTTGTGCC

Rice left targeting Sequence

1/564

JHP

## Figure 34B

pMSK49.seq Length: 5263

```

2751 CATTAAACATC ACCATCTAAT TCACACAGAA TTGGGACAAC TCCAGTGAAA
2801 AGTTCTTCTC CTTTACTagc CATggcgacc tccaattttc cttcaactgc
2851 aagttcTTTG CCCACTACCT TGGTGATCTC GCCTTTCAGC TAGTGGACAA
2901 ATTCTTCCAA CTGATCTGCG CGCGAGGCCA AGCGATCTTC TTCTTGTCCTA
2951 AGATAAGCCT GTCTAGCTTC AAGTATGACG GGCTGATACT GGGCCGGCAG
3001 GCGCTCCATT GCCCAGTCGG CAGCGACATC CTTGGGCGCG ATTTTGCCGG
3051 TTACTGCGCT GTACCAATG CGGGACAACG TAAGCACTAC ATTTGCGTCA
3101 TCGCCAGCCC AGTCGGGCGG CGAGTTCAT AGCGTTAAGG TTTCATTAG
3151 CGCCTCAAAT AGATCCTGTT CAGGAACCGG ATCAAAGAGT TCCTCCGCGG
3201 CIGGACCTAC CAAGGCAACG CTATGTTCTC TTGCTTTTGT CAGCAAGATA
3251 GCCAGATCAA TGTCGATCGT GGCTGGCTCG AAGATACCTG CAAGAATGTC
3301 ATTGCGCTGC CATTCTCCAA ATTGCAGTTC GCGCTTAGCT GGATACGCGC
3351 ACGGAAATGAT GTCGTCGTGC ACAACAATGG TGACTTCTAC AGCGCGGAGA
3401 ATCTCGCTCT CTCCAGGGGA AGCCGAAGTT TCCAAAGGT CGTTGATCAA
3451 AGCTCGCCGC GTTGTTCAT CAAGCCTTAC GGTACCGTA ACCAGCAAAT
3501 CAATATCACT GTGTGGCTTC AGGCCGCCAT CCACTGCGGA GCGTACAAA
3551 TGTACGGCCA GCAACGTCGG TTCGAGATGG CGCTCGATGA CGCCAACTAC
3601 CTCTGATAGT TGAGTCGATA CTTCCGCGAT CACCGCTTCg ctaggcaagt
3651 cttcttcaga aatgagtttt tgttcgctag cTGTCCACC AGTCATGCTT
3701 GCCATATGTA TATCTCCTTC TTAAGTTAA ACAAAATTAT TTCTAGTGGG
3751 AAACCGTTGT GGTCTCCCTC CCAGAAATAT AGCCATCCCT GCCCCTCAC
3801 GTCAATCCCA CGAGCCTCTT ATCCATCTC ATTGAACGAC GGCGGGGGAG
3851 Cgagctcgaa ttctgcage cggatCTAC CATTCCGAA GGAATGGGG
3901 CTACATTTC TTTCAATTC CATTCAAGAG TTTCTTATCT GTTCCACCGC
3951 CCTTTTTTGA GACCTCGAAA CATGAAATGG ACAAAATCCT TCTCTTAGGA
4001 ACACATACAA GAAAAGGAT AATGGTAGCC CTCCTATTAA CTACTTCATT
4051 TCATTTATGA ATTTCATAGT AATAGAAATC CATGTCTTAC CGAGACAGAA
4101 TTTCGAACTT GCTATCCTCT TGCCTAATAG GCAAAGATTG ACCTCTGTAG
4151 AAAGAATGAT TCATTCCGAT CGATATGAGG ACCCAACTAC GTTGCATTGC
4201 AGAATCCATG TTCCATATT GAAGAGGGTT GACCTCTGTG CTTCTCTCAT
4251 GGTACAATCC TCTTCCTGCT GAGCCCCCTT TCTCCTCGGT CCAACAGAAA
4301 AAAATGGAGG ACTGGTGCCG ACAGTTCATC ACGGAAAGAA GAACTCACAG
4351 AGCCGGGATC GCTAACTAAT AGAATAGTAC TACTAATAA TACTAATATA
4401 TAGAAATAGA TATctagcta GAATAGAAA CAACTAATAT ATAGATAATC
4451 GAAATTGAAA AGAACTGTCT TTTCTGTATA CTTTGCCCTT TCTATTGCTA
4501 CCGCGGGTCT TATGCAATCG ATCGGATCAT ATAGATATCC CTTCAACACA
4551 ACATAGGTCA TCGAAAGGAT CTCGGACGAC TCACCAAAGC ACGAAAGCCA
4601 GTTAGAAAT GGATTCTAT TTGAAGAGTG CCTAACCGCA TGGATAAGCT
4651 CACATTAACC CGTCAATTTT GGATCCAATT CGGGATTTT CTTGGGAAGT
4701 TTCGGGAAGA AATTGGAATG GAATAATATA GATTCATACA GAGGAAAAGG
4751 TTCTCTATTG ATGCAAACGC TGTACCTAGA GGATAGGGAT AGAGGAAGAG
4801 GGAAAATCG AAATGAAATA AATAAAGAT AAAGCAAAA AAAAATAAGT
4851 CGAAGATAGA AGAGCCCGA TTCCAAATGA AGAAATGGAA ACTCGAAAAG
4901 GATCCTTCTG ATTCTCAAAG AATGAGGGGC AAGGGGATTG ATACCGAGAA
4951 AGATTTCTTC TTATTATAAG ACGTGATTG ATCCGCATAT GTTTGGTAAA
5001 AGAACAATCT TCTCCTTTAA TCATAAATGG AAAGTGTTC AATTAGAACAT
5051 GAAAACGTGA CTCAATTGGT CTTAGTTAGT CTTGGGACG GAGTGGAGA
5101 AAGGGCGAAG ACTCTCGAAC GAGGAAAAGG ATCCCTTCGA AAGAATTGAA
5151 CGAGGAGCCG TATTAGGTGA AAATCTCATG TACGATTCTG TAGAGGGACA
5201 GGAAGGGTGA CTTATCTGTC GACTTTTCCA CTATCAACCC CAAAAACCC
5251 AACTCTGCCT TAG

```

oada

L7910DB

Rice Right targeting sequence

125878

Gene	Product	Plasmid
<i>aadA16gfp</i>	FLARE16-S	pMSK51 (BS)
<i>aadA16gfp-S1</i>	FLARE16-S1	pMSK56 (Nt-pRV111B)
<i>aadA16gfp-S2</i>	FLARE16-S2	pMSK57 (Nt-pRV111B)
<i>aadA11gfp-S3</i>	FLARE11-S3	pMSK49 (Os-pMSK49)

**Figure 35**



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/17806

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :A01H 4/00, 5/00; C12N 15/29, 15/31, 15/65, 15/67, 15/82, 15/84  
US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EMBL, Genbank

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 97/06250 A1 (RUTGERS UNIVERSITY) 20 February 1997, pages 1, 3-6, 12-14, 17-18 and Figures 1 and 5.	1-6 and 15-17
Y	NIELSEN et al. Characterization of the Pea Chloroplast DNA <i>OriA</i> Region. Plasmid. 1993, Vol. 30, pages 197-211, see page 197.	1-6
Y	SEKIYA et al. Sequence of the gene for isoleucine tRNA <sub>I</sub> and the surrounding region in a ribosomal RNA operon of Escherichia coli. Nucleic Acids Research. 1979, Vol. 6, pages 575-592, see page 575.	1-6



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

07 NOVEMBER 1999

Date of mailing of the international search report

30 NOV 1999

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

DAVID T. FOX

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/17806

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SPRENGART et al. The downstream box: an efficient and independent translation initiation signal in <i>Escherichia coli</i> . The EMBO Journal. 1996, Vol. 15, No. 3, pages 665-674, see page 665.	1-6
Y	VASIL et al. Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. Bio/Technology. June 1992, Vol. 10, No. 6, pages 667-674, see page 667.	7-8 and 18-26
Y	PANG et al. An Improved Green Fluorescent Protein Gene as a Vital Marker in Plants. Plant Physiology. 1996, Vol. 112, pages 893-900, see page 893.	9-14
Y	HIRATSUKA et al. The complete sequence of the rice ( <i>Oryza sativa</i> ) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. Molecular and General Genetics. 1989, Vol. 217, pages 185-194, see page 185.	18-26
Y	ZHANG et al. Efficient regeneration of transgenic plants from rice protoplasts and correctly regulated expression of the foreign gene in plants. Theoretical and Applied Genetics. 1988, Vol. 76, pages 835-840, see page 835.	18-23
Y	THOMPSON et al. Protoplast culture of rice ( <i>Oryza sativa</i> L.) using media solidified with agarose. Plant Science. 1986, Vol. 47, No. 2, pages 123-133, see page 123.	24-26
Y	MAKRIDES, S.C. Strategies for Achieving High-Level Expression of Genes in <i>Escherichia coli</i> . Microbiological Reviews. September 1996, Vol. 60, No. 3, pages 512-538, see page 524.	27-28
Y	MURRAY et al. Codon usage in plant genes. Nucleic Acids Research. 1989, Vol. 17, No. 2, pages 477-498, see pages 477 and 491.	27-28

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/17806

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐

The additional search fees were accompanied by the applicant's protest.

☐

No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/17805

## A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/69.7, 320.1, 424, 430, 430.1, 431, 468, 469, 470; 536/23.6, 23.7, 24.1; 800/278, 287, 288, 292, 293, 294, 300, 300.1, 320, 320.1, 320.2, 320.3

## B. FIELDS SEARCHED

Minimum documentation searched

Classification System: U.S.

435/69.7, 320.1, 424, 430, 430.1, 431, 468, 469, 470; 536/23.6, 23.7, 24.1; 800/278, 287, 288, 292, 293, 294, 300, 300.1, 320, 320.1, 320.2, 320.3

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-26, drawn to a DNA construct comprising a promoter element, a leader, and a downstream box element, and a method for its use for enhanced plastid expression of a heterologous gene.

Group II, claim(s) 27-28, drawn to a method for modifying codon usage in structural genes.

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions of each of Groups I, drawn to a first product and method of use, and Group II, drawn to a second method, are taught or suggested by the prior art, as outlined in the accompanying Search Report, and so each do not constitute an advance over the prior art which would constitute a special technical feature. Furthermore, the various leader sequences and downstream box elements of Group I, and the methods for plastid transformation of Group I, are not required by Group II; while the methods of structural gene modification of Group II are not required by Group I. Accordingly, the inventions do not relate to a single inventive concept which would constitute a single special technical feature.